Intrinsic Stochastic Differential Equations as Jets

J. Armstrong and D. Brigo

Abstract

We explain how Itô Stochastic Differential Equations (SDEs) on manifolds may be defined using 2-jets of smooth functions. We show how this relationship can be interpreted in terms of a convergent numerical scheme. We show how jets can be used to derive graphical representations of Itô SDEs. We show how jets can be used to derive the differential operators associated with SDEs in a coordinate free manner. We relate jets to vector flows, giving a geometric interpretation of the Itô–Stratonovich transformation. We show how percentiles can be used to give an alternative coordinate free interpretation of the coefficients of one dimensional SDEs. We relate this to the jet approach. This allows us to interpret the coefficients of SDEs in terms of "fan diagrams". In particular the median of a SDE solution is associated to the drift of the SDE in Stratonovich form for small times.

1 Introduction

Stochastic Differential Equations (SDEs) on manifolds were first defined by Itô in [25]. Itô's formulation was given in terms of coordinate charts. This has lead many authors to seek coordinate free formulations of SDEs on manifolds. We will describe such a formulation in the language of 2-jets [35]. We will study how this formulation gives rise to intuitive graphical representations of SDEs.

Coordinate free formulations of SDEs have been given previously. One approach is to use Stratonovich calculus (see [11, 10, 34]). Another is the theory of second order tangent vectors, diffusors and Schwartz morphism (see [12, 13]). A third is via the Itô bundle (see [5], [20] or the appendix in [7]).

The value of the 2-jet approach is that it is particularly simple and intuitive. In particular, as the notion of a jet is already familiar to differential geometers, we do not need to introduce novel differential geometric constructs.

In Section 2 we will give an informal description of the definition of SDEs in the language of 2-jets. This description does not require the reader to have prior experience of SDEs (though we do assume they know the definition of Brownian motion). For simplicity we first consider the case of an SDE driven by a single Brownian motion. Our description of SDEs is given by writing down a system of difference equations using a coordinate free notation. A formal proof that the solutions of these equations converge to the solutions of the classically defined Itô SDEs is given in Appendix A.

We also consider how SDEs can be understood graphically. In particular we will see how 2-jets allow us to draw an SDE in a way that makes the transformation law of SDEs, known as Itô's lemma, intuitively clear. We will illustrate a way of drawing an SDE on a rubber sheet such that if the sheet is stretched, the diagram transforms according to Itô's lemma. In other words given an SDE in $\mathbb{R}^n$ we give a method of drawing SDEs such that for all well-behaved $f : \mathbb{R}^n \to \mathbb{R}^n$ the following diagram commutes.

$$\begin{array}{ccc} \text{SDE for X} & \xrightarrow{\text{Itô's lemma}} & \text{SDE for } f(X) \\ \text{Draw}\downarrow & & \text{Draw}\downarrow \\ \text{Picture of SDE for X in } \mathbb{R}^n & \xrightarrow{f} & f(\text{Picture of SDE for X}) \end{array} \tag{1}$$

Moreover, we will show how the language of 2-jets allows us to write a particularly elegant formulation of Itô's lemma.

In Section 3 we describe the relationship between the jet formulation and differential operator formulations of SDEs. We use the language of jets to give geometric expressions for many important concepts that arise in stochastic analysis. These geometric representations are in many ways more elegant than the traditional representations in terms of the coefficients of SDEs. In particular we will give coordinate free formulations of the following: the diffusion operators; Itô SDEs on manifolds and Brownian motion on Riemannian manifolds.

In Section 4, we return to the question of graphical representations of SDEs. We show how to represent processes driven by multiple Brownian motions. We illustrate this using the Heston stochastic volatility model (two-dimensional diffusion) and Brownian motion on the torus.

In Section 5 we consider how our formulation is related to the Stratonovich formulation of SDEs. We will prove that sections of the bundle of n-jets of curves in a manifold correspond naturally to n-tuples of vector fields in the manifold. When translated into a statement about SDEs, the special case when $n = 2$ can be interpreted as the correspondence between Itô calculus and Stratonovich calculus.

In Section 6 we consider an alternative approach to understanding the coefficients of 1-dimensional SDEs based on the coordinate free notion of percentiles. We will see that the 2-jet defining an Itô SDE can be interpreted as defining a *fan diagram* showing the limiting trajectories of certain percentiles of the probability distributions associated with the SDE solution process. Moreover we will show that the drift of the Stratonovich formulation can be similarly interpreted as a short-time approximation to the median. We also consider short time behaviour of the mode.

Our work has a number of applications. Firstly graphical representations of SDEs should be a valuable tool for the qualitative analysis of SDEs and for developing an intuitive understanding of the properties of SDEs. Our illustrations of Itô's lemma give a first example of this. Secondly coordinate free formulations of SDEs will often be considerably simpler than local coordinate formulations and hence should assist in the theoretical development of stochastic differential

geometry. An example of this is given in [2, 3] where a notion of projection for SDEs is defined using the 2-jet approach. It is considerably easier to understand this notion using jets than with a local coordinate formulation. A further application is given in [4] where the jet approach is used to numerically solve SDEs on manifolds. We hope in future work to give applications of this method to statistics similar to those given in [8].

2 SDEs as fields of curves driven by a single Brownian motion

2.1 Drawing and simulating SDEs as "fields of curves"

Suppose that at every point x in $\mathbb{R}^n$ we have an associated smooth curve

$$\gamma_x : \mathbb{R} \to \mathbb{R}^n \quad \text{with } \gamma_x(0) = x.$$

As an example we might define γ^E_x on $\mathbb{R}^2$ as follows

$$\gamma^E_{(x_1,x_2)}(t) = (x_1, x_2) + t(-x_2, x_1) + 3t^2(x_1, x_2).$$

We will use the superscript E to indicate this example curve throughout. This field of curves is plotted in Figure 1.

To be precise we have taken a grid of points in $\mathbb{R}^2$ which are marked as dots in the figure. We have then drawn the curve γ^E_x at each grid point x for the parameter values t in $(-0.1, 0.1)$. In general when drawing such a figure for a general γ, one should use the same range $t \in (-\epsilon, \epsilon)$ for every curve in the figure, but one is free to choose ϵ to make the diagram visually appealing. (In just the same way when drawing vector fields, one chooses a sensible scale for each vector).

As can be seen in the figure, our specific example, γ^E has a circular symmetry. This arises from the radially outward t^2 component and the orthogonal counterclockwise circular component t. Our example has also been chosen to have zero derivatives with respect to t from the third derivative on. This is because we will show how to define a stochastic process in terms of a field of curves γ and we will see that the limiting behaviour of this process only depends upon the first and second order terms in t.

Given such a γ, a starting point X_0 (the deterministic $x_0 = (1, 0)$ in our example, $X_0 = x_0$), a Brownian motion W_t and a time step δt we can define a discrete time stochastic process, $X^{\delta t}$ using the following recurrence relation

$$X^{\delta t}_0 := x_0, \quad X^{\delta t}_{t+\delta t} := \gamma_{X^{\delta t}_t}(W_{t+\delta t} - W_t). \tag{2}$$

In Figure 2 we have plotted the trajectories of the process for γ^E, the starting point $(1, 0)$, a fixed realization of Brownian motion and a number of different time steps. Rather than just plotting a discrete set of points for this discrete time process, we have connected the points using the curves in $\gamma^E_{X_t}$.

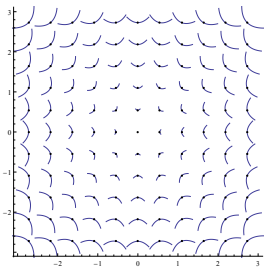

Figure 1: A plot of γ^E

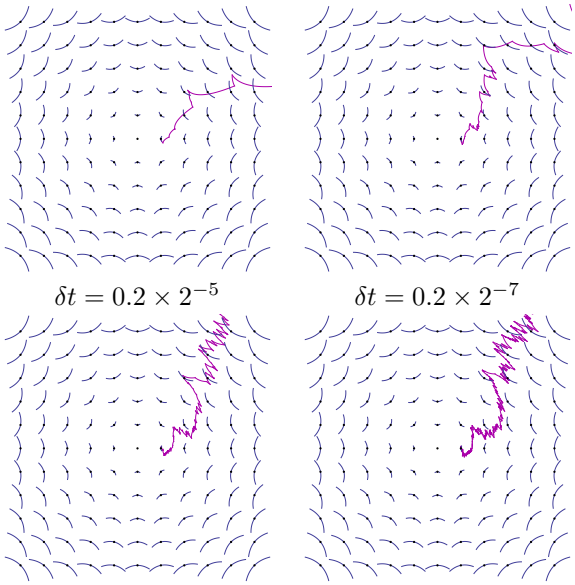

Figure 2: Discrete time trajectories for γ^E for a fixed W_t and X_0 with different values for δt

Notice that since the $\delta W_t = W_{t+\delta t} - W_t$ are normally distributed with standard deviation $\sqrt{\delta t}$ we can interpret the trajectories as being randomly generated trajectories that move from $X_t^{\delta t}$ to $X_{t+\delta t}^{\delta t}$ by following the curve $s \mapsto \gamma_{X_t^{\delta t}}(s)$ from $s = 0$ to $s = \epsilon_t \sqrt{\delta t}$ where the ϵ_t are independent normally distributed random variables.

As the figure suggests, these discrete time stochastic processes (2) converge in some sense to a limit as the time step tends to zero.

We will use the following notation for the limiting process.

$$\text{Coordinate free SDE:} \quad X_t \curvearrowright \gamma_{X_t}(\mathrm{d}W_t), \quad X_0 = x_0. \tag{3}$$

For the time being, let us simply treat equation (3) as a short-hand way of saying that equation (2) converges in some sense to a limit. Note that it will not converge for arbitrary γ's but it does converge for nice γ such as γ^E or γ's with sufficiently good regularity. The reader familiar with Itô calculus will want to know how this notation corresponds to Itô stochastic differential equations and in precisely what sense and under what circumstances equation (2) converges to a limit. These questions are addressed in Section 2.2.

An important feature of equation (2) is that it makes no reference to the vector space structure of $\mathbb{R}^n$ for our state space X. We have maintained this in the formal notation used in equation (3). By avoiding using the vector space structure on $\mathbb{R}^n$ we will be able to obtain a coordinate free understanding of stochastic differential equations.

Example 2.1. For a fixed $\alpha \in \mathbb{N}$, in a given coordinate system on $\mathbb{R}$, we can define curves at each point in $\mathbb{R}$ by:

$$\gamma_x^\alpha(s) = x + s^\alpha$$

Let us compute the limit of the discrete time process corresponding to these curves. In the case $\alpha = 1$, we have trivially that the $X_t = x_0 + W_t$. By equation (2) we have that

$$\begin{aligned} X_{n\delta t}^{\delta t} &= x_0 + \sum_{i=1}^{n} (W_{(i+1)\delta t} - W_{i\delta t})^\alpha \\ &= x_0 + (\delta t)^{\frac{\alpha}{2}} \sum_{i=1}^{n} \epsilon_i^\alpha \end{aligned}$$

where the ϵ_i are independently normally distributed with mean 0 and standard deviation 1. Fixing a terminal time T so that $\delta t = \frac{T}{n}$ we have

$$X_T^{\delta t} = x_0 + (T/n)^{\frac{\alpha}{2}} \sum_{i=1}^{n} \epsilon_i^\alpha.$$

By the strong law of large numbers we see that if $\alpha = 2$ this converges a.s. to $x_0 + T$. If $\alpha \geq 3$ we see that this converges a.s. to x_0.

2.2 SDEs as fields of curves up to order 2: 2-jets

Let us now invoke explicitly the $\mathbb{R}^n$ structure of the state space by choosing a specific coordinate system and consider the (component-wise) Taylor expansion of γ_x. We have

$$\gamma_x(t) = x + \gamma_x'(0)t + \frac{1}{2}\gamma_x''(0)t^2 + R_x t^3, \quad R_x = \frac{1}{6}\gamma_x'''(\xi), \quad \xi \in [0, t],$$

where $R_x t^3$ is the remainder term in Lagrange form. Substituting this Taylor expansion in our Equation (2) we obtain

$$\delta X_t^{\delta t} = \gamma'_{X_t^{\delta t}}(0)\delta W_t + \frac{1}{2}\gamma''_{X_t^{\delta t}}(0)(\delta W_t)^2 + R_{X_t^{\delta t}}(\delta W_t)^3, \quad X_0^{\delta t} = x_0. \tag{4}$$

Example 2.1 suggests that we can replace the term $(\delta W_t)^2$ with δt and we can ignore terms of order $(\delta W_t)^3$ and above. So we expect that under reasonable conditions, in the chosen coordinate system, the recurrence relation given by (2) will converge to the same limit as the numerical scheme

$$\delta \bar{X}_t^{\delta t} = \gamma'_{\bar{X}_t^{\delta t}}(0)\delta W_t + \frac{1}{2}\gamma''_{\bar{X}_t^{\delta t}}(0)\delta t, \quad \bar{X}_0^{\delta t} = x_0.$$

Defining $a(X) := \gamma_X''(0)/2$ and $b(X) := \gamma_X'(0)$ we have that this last equation can be written as

$$\delta \bar{X}_t^{\delta t} = a(\bar{X}_t^{\delta t})\delta t + b(\bar{X}_t^{\delta t})\delta W_t. \tag{5}$$

It is well known that this last scheme (Euler scheme) does converge in some appropriate sense to a limit ([28]). This limit is more conventionally written as the solution to the Itô stochastic differential equation

$$\mathrm{d}\tilde{X}_t = a(\tilde{X}_t)\,\mathrm{d}t + b(\tilde{X}_t)\mathrm{d}W_t, \quad \tilde{X}_0 = x_0. \tag{6}$$

The coefficient $a(\tilde{X}_t)$ is often referred to as the *drift*. The coefficient $b(\tilde{X}_t)$ is often referred to as the *diffusion coefficient* (also known as volatility in applications to social sciences). Thus, given a coordinate system, we may think of equation (2) as defining a numerical scheme for approximating the Itô SDE (6). In this context we call (2) the 2-jet scheme. A rigorous proof of the convergence of the 2-jet scheme in mean square ($L^2(\mathbb{P})$) to the solution of the Itô SDE, based on appropriate bounds on the derivatives of the curves γ_x, is given in Appendix A. This notion of convergence is not fully coordinate independent, however in Appendix B we describe a fully coordinate free notion of convergence which we call mean square convergence on compacts. Our proof of convergence in $L^2(\mathbb{P})$ imples that the 2-jet scheme will always converge in mean square on compacts if the coefficients are sufficiently smooth.

At this point one may wonder in which sense Equation (2) and its limit are coordinate free. It is important to note that the coefficients of equation (6) only depend upon the first two derivatives of γ. We say that two smooth curves $\gamma : \mathbb{R} \to \mathbb{R}^n$ have the same k-jet ($k \in \mathbb{N}, k > 0$) if they are equal up to order $O(t^k)$ in a given coordinate system. If this holds in a given coordinate system, it will hold in all coordinate systems. More generally we have:

Definition 2.1. A k-jet of a function between smooth manifolds M and N is defined to be the equivalence class of all smooth maps $f : M \to N$ that are equal up to order k in one, and hence all, coordinate systems.

Using this terminology, we say that the coefficients of equation (6) (and (5)) are determined by the 2-jet of a curve $\gamma : \mathbb{R} \to \mathbb{R}^n$ in a specific coordinate system.

In the light of the above convergence result, we can say that in pictures such as Figure 1 one should avoid interpreting any details other than the first two derivatives of the curve. One way of doing this is by insisting that we draw the quadratic curves that best fit the curves γ rather than the actual curve γ itself.

Notice that vectors can be defined in the same way as 1-jets of smooth curves. In just the same way as we draw quadratic curves in Figure 1, one normally chooses to draw straight lines in a diagram of a vector field.

In summary: vector fields are fields of 1-jets and they represent ODE's; diagrams such as Figure 1 are pictures of fields of 2-jets and they represent Itô SDEs.

Given a curve γ_x, we will write $j_2(\gamma_x)$ for the two jet associated with γ_x. This is formally defined to be the equivalence class of all curves which are equal to γ_x up to $O(t^2)$ included.

Since we will show that, under reasonable regularity conditions, the limit of the symbolic equation (3) depends only on the 2-jet of the driving curve, we

may rewrite equation (3) as

$$\text{Coordinate-free 2-jet SDE:} \qquad X_t \curvearrowright j_2(\gamma_{X_t})(\mathrm{d}W_t), \quad X_0 = x_0. \tag{7}$$

This may be interpreted either as a coordinate free notation for the classical Itô SDE given by equation (6) or as a shorthand notation for the limit of the process given by the discrete time equation (2).

2.3 Coordinate-free Itô formula with jets

Suppose that f is a smooth mapping from $\mathbb{R}^n$ to itself and suppose that X satisfies (2). It follows that $f(X)$ satisfies

$$(f(X^{\delta t}))_{t+\delta t} = f \circ \gamma_{X_t^{\delta t}}(\delta W_t).$$

Taking the limit as δt tends to zero we have:

Lemma 2.1. *[Itô's lemma — coordinate free formulation] If the process X_t satisfies*

$$X_t \curvearrowright j_2(\gamma_{X_t})(\mathrm{d}W_t)$$

then, writing $\times$ for the Cartesian product of functions, $(X_t, f(X_t))$ satisfies

$$(X_t, f(X)_t) \curvearrowright j_2((\gamma_{X_t} \times f \circ \gamma_{X_t}))(\mathrm{d}W_t)$$

(we might also write more directly, with abuse of notation, $f(X)_t \curvearrowright j_2(f \circ \gamma_{X_t})(\mathrm{d}W_t)$).

If one prefers the more traditional format for SDEs given in (6) we simply need to calculate the derivatives of $f \circ \gamma$. In a chosen coordinate system, let us write γ_X^i for the i-th component of γ_X with respect to the coordinates $x_1, x_2, \ldots x_n$ for $\mathbb{R}^n$. Two applications of the chain rule give

$$\begin{aligned}
(f \circ \gamma_X)'(t) &= \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(\gamma_X(t)) \frac{\mathrm{d}\gamma_X}{\mathrm{d}t} \\
(f \circ \gamma_X)''(t) &= \sum_{j=1}^{n} \sum_{i=1}^{n} \frac{\partial^2 f}{\partial x_i \partial x_j}(\gamma_X(t)) \frac{\mathrm{d}\gamma_X^i}{\mathrm{d}t} \frac{\mathrm{d}\gamma_X^j}{\mathrm{d}t} \\
&\quad + \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(\gamma_X(t)) \frac{\mathrm{d}^2\gamma_X}{\mathrm{d}t^2}
\end{aligned}$$

We conclude that our lemma is equivalent to the classical Itô's lemma.

We can now interpret Itô's lemma geometrically as the statement that the transformation rule for jets under a change of coordinates is given by composition of functions.

Since we now understand the geometric content of Itô's lemma, we can draw a picture to illustrate it. Consider the transformation $\phi : \mathbb{R}^2/\{0\} \to [-\pi, \pi] \times \mathbb{R}$ by

$$(\theta, s) = \phi(x_1, x_2) = \left(\arctan(x_2/x_1), \log(\sqrt{x_1^2 + x_2^2}) \right),$$

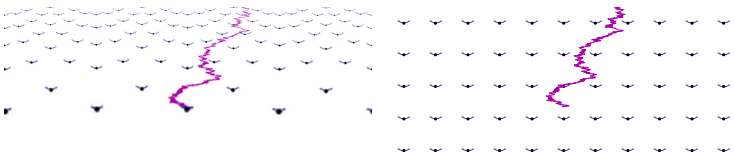

Figure 3: Two plots of the process $j_2(\phi \circ \gamma^E)$ in the plane (θ, s). The left plot was generated by transforming the image in Cartesian coordinates pixel by pixel. The right was computed using Itô's lemma.

or equivalently

$$\phi(\exp(s)\cos(\theta), \exp(s)\sin(\theta)) = (\theta, s),$$

applied to our example process γ^E. This can be viewed as a transformation of the complex plane $\phi(z) = i\log(z)$. We use ϕ to transform the bottom right picture in Figure 2 in two different ways. Firstly we apply directly ϕ to each point of Figure 2 to obtain a new point to be inserted in a new figure. This is done by using image manipulation software. In other words we stretch the image without any consideration of its mathematical structure. The result of this is shown in the left hand side of Figure 3.

As an alternative approach, we transform our equation using Itô's lemma applied to the function ϕ. So equation (8) below for (X_1, X_2) transforms to the equation (9) for (θ, s).

$$d(X_1, X_2) = 3(X_1, X_2)\,dt + (-X_2, X_1)\,dW_t, \tag{8}$$

$$d(\theta, s) = 3\left(0, \frac{7}{2}\right)dt + (1, 0)\,dW_t. \tag{9}$$

We can then use this equation to plot the process (θ, s) directly by simulating the process in discrete time as before. The result is shown in the right hand side of Figure 3.

As one can see the two approaches to plotting the transformed process give essentially identical results, showing an example of our earlier diagram (1) at work. The differences one can see are: the lower quality in the left image, obtained by transforming pixels rather than using vector graphics; the grid points at which the 2-jets are plotted are changed; small differences in the simulated path since we have only simulated discrete time paths.

We have assumed that the 2-jet $j_2(\gamma_x)$ is associated in a deterministic and time independent manner with the point x. However, we are investigating how the theory can be generalized to time dependent and stochastic choices of 2-jets.

2.4 SDEs driven by vector-Brownian motion as 2-jets

Consider jets of functions of the form

$$\gamma_x : \mathbb{R}^d \to \mathbb{R}^n.$$

Just as before we can consider discrete time difference equations of the form

$$X_{t+\delta t} := \gamma_{X_t}\left(\delta W_t^1, \ldots, \delta W_t^d\right), \tag{10}$$

or, if writing $\delta W_t^\alpha = \epsilon_\alpha \sqrt{\delta t}$, with ϵ independent normals,

$$X_{t+\delta t} := \gamma_{X_t}\left(\epsilon_1\sqrt{\delta t}, \ldots, \epsilon_d\sqrt{\delta t}\right).$$

Again, the limiting behaviour of such difference equations will only depend upon the 2-jet $j_2(\gamma_x)$ and can be denoted by (7), where it is now understood that dW_t is the vector Brownian motion increment.

The multi-dimensional analogue of Example 2.1 suggests we write $\delta W^\alpha \delta W^\beta \approx g_E^{\alpha\beta}\delta t$. Here $g_E^{\alpha\beta}$ denotes the Euclidean metric on $\mathbb{R}^d$. Thus $g_E^{\alpha\beta}$ is equal to 1 if α equals β and 0 otherwise. We can now compute a second order Taylor expansion as follows.

$$\delta X_t \approx \frac{1}{2}\sum_\alpha \sum_\beta \frac{\partial^2 \gamma_{X_t}}{\partial u^\alpha \partial u^\beta} g_E^{\alpha\beta}\delta t + \sum_\alpha \frac{\partial \gamma_{X_t}}{\partial u^\alpha}\delta W_t^\alpha. \tag{11}$$

Here u^α are the standard orthonormal coordinates for $\mathbb{R}^d$. We have chosen to write $g_E^{\alpha\beta}$ instead of using a Kronecker δ because one might want to choose non-orthonormal coordinates for $\mathbb{R}^d$ and so it is useful to notice that g_E transforms as a symmetric 2-form and not an endomorphism. Another advantage is that we can use the Einstein summation convention. For example (11) can be rewritten as

$$\delta X_t^i = \frac{1}{2}\partial_\alpha \partial_\beta \gamma^i g_E^{\alpha\beta}\delta t + \partial_\alpha \gamma^i\, \delta W_t^\alpha. \tag{12}$$

We have shown informally how to define an SDE as the limit of a numerical scheme defined in terms of 2-jets and we have shown how this scheme can be written in local coordinates. A reader who is familiar with classical Itô calculus will immediately recognize (12) as the Euler scheme for the Itô stochastic differential equation obtained by replacing each δ in (12) with a d. We now state the relationship between these two approaches formally.

Theorem 2.1. Convergence of the 2-jet schemes to Itô SDEs. *Let $\gamma_x : \mathbb{R}^d \to \mathbb{R}^n$ be a smoothly varying family of functions whose first and second derivatives in $\mathbb{R}^d$ satisfy Lipschitz conditions (and hence linear growth bounds). In other words we require that there exists a positive constant K such that for*

all $x, y \in \mathbb{R}^n$ *and* $\alpha, \beta \in \{1, 2, \ldots, d\}$ *we have*

$$\left| \frac{\partial \gamma_x}{\partial u^\alpha}\Big|_{u=0} - \frac{\partial \gamma_y}{\partial u^\alpha}\Big|_{u=0} \right| \leq K|x-y|,$$

$$\left| \frac{\partial^2 \gamma_x}{\partial u^\alpha \partial u^\beta}\Big|_{u=0} - \frac{\partial \gamma_y}{\partial u^\alpha \partial u^\beta}\Big|_{u=0} \right| \leq K|x-y|,$$

$$\left(\textit{and hence } \left| \frac{\partial \gamma_x}{\partial u^\alpha}\Big|_{u=0} \right|^2 \leq K^2 \left(1 + |x|^2\right), \right.$$

$$\left. \left| \frac{\partial^2 \gamma_x}{\partial u^\alpha \partial u^\beta}\Big|_{u=0} \right|^2 \leq K^2 \left(1 + |x|^2\right) \quad \right).$$

Note that we are using the letter u *to denote the standard orthonormal coordinates for* $\mathbb{R}^d$*. Suppose in addition that we have a uniform bound on the third derivatives:*

$$\left| \frac{\partial^3 \gamma_x}{\partial u^\alpha \partial u^\beta \partial u^\gamma}\Big|_{u=0} \right| \leq K.$$

Let T *be a fixed time and let* $\mathcal{T}^N := \{0, \delta t, 2\,\delta t, \ldots, N\,\delta t = T\}$*, be a set of discrete time points. Let* $X_t^{\delta t}$ *(we will omit the superscript below) denote the* 2*-jet scheme defined by*

$$X_{t+\epsilon} = \gamma_{X_t}\left(\frac{\epsilon}{\delta t}(W_{t+\delta t} - W_t)\right), \qquad t \in \mathcal{T}^{N-1},\ \epsilon \in [0, \delta t]\ , \qquad X_0 = x_0. \tag{13}$$

This converges in $L^2(\mathbb{P})$ *to the classical Itô solution of the corresponding SDE, namely*

$$\tilde{X}_t = \tilde{X}_0 + \int_0^t a(\tilde{X}_s)\,\mathrm{d}s + \sum_{\alpha=1}^d \int_0^t b_\alpha(\tilde{X}_s)\,\mathrm{d}W_s^\alpha, \quad t \in [0, T] \tag{14}$$

where

$$a(x) := \frac{1}{2}\sum_{\alpha=1}^d \frac{\partial^2 \gamma_x}{\partial u^\alpha \partial u^\alpha}\Big|_{u=0}$$

and

$$b_\alpha(x) := \frac{\partial \gamma_x}{\partial u^\alpha}\Big|_{u=0}.$$

More precisely we have the estimate

$$\sup_{t\in[0,T]} E\left\{|X_t^N - \tilde{X}_t|^2\right\} \leq C\ \delta t = C\frac{T}{N} \tag{15}$$

for some constant C *independent of* N*. We denote the coordinate free equation obtained as limit of* (13) *by*

$$X_t \curvearrowright j_2(\gamma_{X_t})(dW_t).$$

This theorem proves one of the main results of this paper: a Itô SDE can be represented in a coordinate free manner simply as a 2-jet driven by Brownian motion. The proof is given in Appendix A.

3 Jets and second order operators

Definition 3.1. (Coordinate free Itô SDEs driven by vector Brownian motion). An Itô SDE or Itô diffusion on a manifold M is a section of the bundle of 2-jets of maps $\mathbb{R}^d \to M$ together with d Brownian motions W_t^i, $i = 1, \ldots, d$.

Our discrete time formulation (equation (10)) of an Itô SDE on a manifold is already coordinate free in that it makes no mention of the vector space structure of $\mathbb{R}^n$. Given such a γ we can write down a corresponding Itô SDE using the notation of equation (7). This may be interpreted either as indicating the limit of the discrete time process or as a short-hand for a classical Itô SDE. The reformulation of Itô's lemma in the language of jets shows that this second interpretation will be independent of the choice of coordinates. The only issue one needs to consider are the bounds needed to ensure existence of solutions. The details of transferring the theory of existence and uniqueness of solutions of SDEs to manifolds are considered in, for example, [11], [10], [25], [12] and [23]. The point we wish to emphasize is that the coordinate free formulation of an SDE given in equation (7) can be interpreted just as easily on a manifold as on $\mathbb{R}^n$.

We will now see how this allows us to study the differential geometry of SDEs in a coordinate free manner. In particular we will show how to give coordinate free definitions of key concepts such as the diffusion operators and Brownian motion.

Note that an alternative approach to stochastic differential geometry is to place these operators centre stage. This is the essence of the approach of second order tangent vectors and Schwarz morphisms. Thus this section can also be seen as establishing the relationship between these approaches.

Suppose that f is a function mapping M to $\mathbb{R}$. We can define a differential operator acting on *functions* in terms of a 2-jet associated with γ_x as follows.

Definition 3.2. (Backward diffusion operator via 2-jets). The Backward diffusion operator for the Itô SDE corresponding to γ_x is defined on suitable functions f as

$$\mathcal{L}_{\gamma_x} f := \frac{1}{2}\Delta_E(f \circ \gamma_x)\Big|_0 = \frac{1}{2}\partial_\alpha \partial_\beta (f \circ \gamma_x)\Big|_0 g_E^{\alpha\beta}. \tag{16}$$

Here Δ_E is the Laplacian defined on $\mathbb{R}^d$. $\mathcal{L}_{\gamma_x}$ acts on functions defined on M.

In contexts where the SDE is understood we will simply write $\mathcal{L}$.

Note that this definition is simply the drift term of the Itô SDE for $f(X)$ computed using Itô's lemma. To first order, the drift measures how the expectation of an SDE solution process changes over time. Thus, with δ denoting the forward t increment as usual,

$$(\delta \mathbb{E}[f(X_t)|X_s = x])\,|_{s=t} = (\mathcal{L}_{\gamma_x}(f))\delta t + O(\delta t^2). \tag{17}$$

Before proceeding to define the forward diffusion operator, let us briefly recall the theory of the bundle of densities on a manifold.

Recall that given a vector space V and a group homomorphism

$$\tau : GL(n, \mathbb{R}) \to \mathrm{Aut}(V)$$

we can define an associated bundle $\mathbf{V}$ over M. The fibres of the bundle over a point p are given by equivalence classes of charts $\phi : M \to \mathbb{R}^n$ and vectors $v \in V$ under the relation

$$(\phi_1, v_1) \sim (\phi_2, v_2) \quad \Leftrightarrow \quad \tau\left((\phi_1)_* \circ (\phi_2)_*^{-1}\right)_p (v_2) = v_1.$$

Each chart $\phi : M \to \mathbb{R}^n$ defines local coordinates on $\mathbf{V}$. We use this to define the smooth structure on $\mathbf{V}$. This generalizes straightforwardly [29] to allow one to associate a vector bundle with any principal G-bundle and a representation of the Lie group G. In our special case, the principal G-bundle we are using is the frame bundle of the manifold.

Now consider the representation

$$\tau(g) = |\det(g)| \in \mathrm{Aut}(\mathbb{R}).$$

This defines a bundle over a manifold M called the bundle of densities. This bundle is denoted Vol. The usual transformation formula for probability densities under changes of coordinates tells us that a probability density over M is a section of Vol.

Integration defines a pairing between functions and densities on M by

$$\int : \quad \Gamma(\mathbb{R}) \times \Gamma(\mathrm{Vol}) \to \mathbb{R}$$
$$\text{by } \int (f, \rho) := \int f\rho.$$

On non-compact manifolds one must either insist that one of f or ρ is compactly supported or consider decay rates of f and ρ to ensure this is well defined.

Note that from a probabilistic point of view, this pairing is interpreted as taking expectations.

Using integration by parts, we can define $\mathcal{L}^*$ to be the formal adjoint of $\mathcal{L}$ with respect to this pairing. This is called the *forward diffusion operator*.

Even if one does not know the initial state X_0 but knows its probability density, ρ, one may integrate equation (17) to obtain

$$\frac{\partial}{\partial t} \int_M f(\rho) = \int_M (\mathcal{L}f)(\rho).$$

Since this holds for all smooth compactly supported functions f we can deduce

$$\frac{\partial \rho}{\partial t} = \mathcal{L}^*(\rho).$$

We conclude that the Fokker–Planck equation follows from Itô's lemma for functions.

Notice that both $\mathcal{L}$ and $\mathcal{L}^*$ are linear second order operators. The key difference is that they have different domains. $\mathcal{L}$ acts on functions; $\mathcal{L}^*$ acts on densities. This gives a geometric explanation as to why $\mathcal{L}$ appears in the Feynman–Kac equation which tells us about the evolution of expectations of functions whereas $\mathcal{L}^*$ appears in the Fokker–Planck equation which tells us about the evolution of probability densities.

3.1 Weak and strong equivalence

We see that both the Itô SDE (12) and the backward diffusion operator (16) use only part of the information contained in the 2-jet. Specifically only the diagonal terms of $\partial_\alpha \partial_\beta \gamma^i$ (those with $\alpha = \beta$) influence the SDE and even for these terms it is only their average value that is important. The same consideration applies to the backward diffusion operator. This motivates the following

Definition 3.3. We say that two 2-jets γ_x^1 and γ_x^2

$$\gamma_x^i : \mathbb{R}^d \to M$$

are *weakly equivalent* if

$$\mathcal{L}_{\gamma_x^1} = \mathcal{L}_{\gamma_x^2}.$$

We say that γ^1 and γ^2 are *strongly equivalent* if in addition

$$j_1(\gamma^1) = j_1(\gamma^2).$$

We see that the SDEs defined by the two sets of 2-jets are equivalent if the 2-jets are strongly equivalent. This means that given the same realization of the driving Brownian motions W_t^α the solutions of the SDEs will be almost surely the same (under reasonable assumptions to ensure pathwise uniqueness of solutions to the SDEs).

When the 2-jets are weakly equivalent, the transition probability distributions resulting from the dynamics of the related SDEs are the same even though the dynamics may be different for any specific realisation of the Brownian motions. For this reason one can define a diffusion process on a manifold as a smooth selection of a second order linear operator $\mathcal{L}$ at each point that determines the transition of densities. In this context $\mathcal{L}$ is known as the *infinitesimal generator of the diffusion*. A diffusion can be realised locally as an SDE, but not necessarily globally.

Recall that the top order term of a quasi linear differential operator is called its symbol. In the case of a second order quasi linear differential operator D which maps $\mathbb{R}$-valued functions to $\mathbb{R}$-valued functions, the symbol defines a section of S^2T, the bundle of symmetric tensor products of tangent vectors, which we will call g_D.

In local coordinates, if the top order term of D is

$$Df = a^{ij}\partial_i \partial_j f + \text{lower order}$$

then g_D is given by

$$g_D(X_i, X_j) = a^{ij} X_i X_j.$$

We are using the letter g to denote the symbol for a second order operator because, in the event that g is positive definite and $d = \dim M$, g defines a Riemannian metric on M. In these circumstances we will say that the SDE/diffusion is *non-singular*. Thus we can associate a canonical Riemannian metric $g_{\mathcal{L}}$ to any non-singular SDE/diffusion.

Definition 3.4. A diffusion on a manifold M is called a Riemannian Brownian motion if

$$\mathcal{L}(f) = \frac{1}{2}\Delta_{g_{\mathcal{L}}}(f).$$

Note that given a Riemannian metric h on M there is a unique Riemannian Brownian motion (up to diffusion equivalence) with $g_{\mathcal{L}} = h$. This is easily checked with a coordinate calculation.

This completes our definitions of the key concepts in stochastic differential geometry and indicates some of the important connections between stochastic differential equations, Riemannian manifolds, second order linear elliptic operators and harmonic maps.

3.2 Brownian Motion

Let us examine the important special case of Brownian motion from a variety of perspectives: local coordinates; the exponential map; and mean curvature.

3.2.1 Local coordinates

We will now show how the definitions above allow us to compute Brownian motion on a topologically non-trivial d-dimensional Riemannian manifold (M, g) for which we have an atlas. We used this computation to generate Figure 6 where we simulate Brownian motion on a genus 2 surface.

Let $(x^1, x^2, \ldots, x^d)$ be a chart for M and let $(u^1, u^2, \ldots, u^d)$ be the standard coordinates on $\mathbb{R}^d$ (so $g_E^{\alpha,\beta}$ is the identity matrix). Let $\gamma_x : \mathbb{R}^d \to M$ with $\gamma_x(0) = x$ and let $f : M \to \mathbb{R}$. Let us write γ^i for the components of γ, so $\gamma^i = x^i \circ \gamma$. We compute

$$\begin{aligned}\mathcal{L}_{\gamma_x} f &= \frac{1}{2} g_E^{\alpha\beta} \frac{\partial^2}{\partial u^\alpha \partial u^\beta}(f \circ \gamma_x)\Big|_{u=0} \\ &= \left(\frac{1}{2} g_E^{\alpha\beta} \frac{\partial^2 f}{\partial x^i \partial x^j}(\gamma_x(u)) \frac{\partial \gamma_x^i}{\partial u^\alpha} \frac{\partial \gamma_x^j}{\partial u^\beta}\right)\Bigg|_{u=0} + \left(\frac{1}{2} g_E^{\alpha\beta} \frac{\partial f}{\partial x^i}(\gamma_x(u)) \frac{\partial^2 \gamma_x^i}{\partial u^\alpha \partial u^\beta}\right)\Bigg|_{u=0}.\end{aligned}$$

Let us assume that γ_x is a quadratic function of u so that in local coordinates

$$\gamma_x^i = x + b_\alpha^i u^\alpha + a_{\alpha\beta}^i u^\alpha u^\beta$$

where b^i and $a^i_{\alpha\beta}$ are real valued functions on the manifold and $i, \alpha, \beta \in \{1, \dots d\}$. So

$$\mathcal{L}_{\gamma_x} f = \frac{1}{2} g_E^{\alpha\beta} b^i_\alpha b^j_\beta \frac{\partial^2 f}{\partial x^i \partial x^j} + g_E^{\alpha\beta} a^i_{\alpha\beta} \frac{\partial f}{\partial x^i}.$$

Following standard conventions, on a Riemannian manifold we write g_{ij} for the metric tensor in local coordinates, we write $|g|$ as an abbreviation for $\det g_{ij}$ and we write g^{ij} for the inverse matrix of g_{ij}. The Laplacian on (M, g) is then given in local coordinates by

$$\begin{aligned}\Delta f &= \frac{1}{\sqrt{|g|}} \frac{\partial}{\partial x^i} \left(\sqrt{|g|} g^{ij} \frac{\partial f}{\partial x^j} \right) \\ &= \frac{1}{\sqrt{|g|}} \frac{\partial}{\partial x^i} \left(\sqrt{|g|} g^{ij} \right) \frac{\partial f}{\partial x^j} + g^{ij} \frac{\partial^2 f}{\partial x^i \partial x^j},\end{aligned}$$

(see, for example, [26]). We have said that γ defines Brownian motion on (M, g) if $\mathcal{L}_{\gamma_x}$ is equal to half the Laplacian operator. So γ will define Brownian motion if at each point x,

$$g_E^{\alpha\beta} b^i_\alpha b^j_\beta = g^{ij} \tag{18}$$

$$g_E^{\alpha\beta} a^i_{\alpha\beta} = \frac{1}{2\sqrt{|g|}} \frac{\partial}{\partial x^j} \left(\sqrt{|g|} g^{ij} \right). \tag{19}$$

As we would expect, these equations are under-determined. We can find a solution to (18) by taking the matrix b^i_α to be the Cholesky decomposition of the matrix g^{ij}. We can also find a solution of equation (19) by taking

$$a^i_{\alpha\beta} = \begin{cases} \frac{1}{2d\sqrt{|g|}} \frac{\partial}{\partial x^j} (\sqrt{|g|} g^{ij}) & \text{if } \alpha = \beta \\ 0 & \text{otherwise.} \end{cases}$$

In summary we have found a canonical choice of γ that locally defines Brownian motion in a chart. Given an atlas we can then choose γ_x at each point by choosing a γ_x from one of the charts around x. Although γ_x itself will not vary smoothly between charts, the weak equivalence class of γ_x will vary smoothly.

In Figure 6 we show the result of simulating Brownian motion on the genus 2 surface in $\mathbb{R}^3$ given in coordinates (y_1, y_2, y_3) by

$$((y_1 - 1)y_1^2(y_1 + 1) + y_2^2)^2 + y_3^2 = \frac{1}{30}.$$

We found 14 charts for this surface by projecting along each of the axes y_i (Mathematica's Solve function made this easy to do). At each point x in this manifold, we chose a specific one of these charts containing x by projecting along the axis whose inner product with the normal at x had the largest absolute value (we preferred the axis y_i with the lowest index i in the event of a tie). In this way we were able to define an explicit quadratic map γ_x at each point. Since the image of $\gamma_x(u)$ will leave the chart for large u, we defined a new function by $\tilde{\gamma}(x) = \gamma_x(u)\rho(|\gamma_x(u)|)$ where $\rho : [0, \infty] \to [0, \epsilon)$ is a smooth increasing function

equal to the identity near 0 and where the value ϵ was so as to ensure that $\tilde{\gamma}$ would never leave the selected chart around x. By construction $j_2(\tilde{\gamma}) = j_2(\gamma)$, so the maps $\tilde{\gamma}$ can be used to approximate Brownian motion in discrete time using equation (10).

3.2.2 Exponential map

Another choice of canonical map $\gamma_x : \mathbb{R}^d \to M$ that generates Brownian motion on a Riemannian manifold is the exponential map [26]. In this case the convergence of the discrete time scheme (10) is well-known. In those situations where the exponential map can be calculated explicitly this gives the most obvious choice of γ. Our local coordinate calculation shows how Brownian motion can be simulated when the exponential map is not explicitly known.

We note that simulating Brownian motion on a Riemannian manifold provides a useful tool for sampling from probability distributions on that manifold. For example, [8] uses the exponential map to simulate Brownian motion and provides explicit coordinate calculations in the case of Stiefel manifolds and show with examples how this can be applied to statistical problems such as dimension-reduction. The papers [19, 9, 6] discuss related approaches to sampling from manifolds.

One does not need a metric to define the exponential map, it can be defined using a connection alone. The scheme (10) resulting from the exponential map of a connection was studied in [18] and a version of Theorem 2.1 was proved for this case (see also the similar paper [33]). This approach to defining SDEs on manifolds is known as the McKean–Gangolli injection.

3.2.3 Mean Curvature

Brownian motion on a hypersurface H in $\mathbb{R}^{d+1}$ can be defined using a d-dimensional stochastic process in $\mathbb{R}^{d+1}$ that has the property that trajectories which start on H stay on H. Let us see how this can be understood geometrically with jets.

Suppose that we have a map $\gamma_x : \mathbb{R}^d \to \mathbb{R}^{d+1}$ at each point x of H. Choose orthonormal coordinates u^α for $\mathbb{R}^d$. For each $\alpha = 1 \dots d$, consider the curves $\gamma_x^\alpha : \mathbb{R} \to \mathbb{R}^{d+1}$ by $\gamma_x^\alpha = \gamma \circ i^\alpha$ where $i^\alpha : \mathbb{R} \to \mathbb{R}^d$ is the inclusion given by coordinate α. If γ_x were the exponential map, each γ_x^α would be parametrised by arc-length and would have curvature orthogonal to H. Moreover the mean of these curvatures as α varies from 1 to d would be the mean curvature vector of H. In coordinates, if the γ_x^α are parameterized by arc-length the mean of the curvatures is given by

$$\frac{1}{d} g_E^{\alpha\beta} \partial_\alpha \partial_\beta \gamma_x^i.$$

We note that this quantity is not dependent on the choice of coordinates u^α and is equal to $\frac{2}{d}$ times the drift term in (12).

From our discussion of weak equivalence of 2-jets, we deduce that γ_x defines Brownian motion on a hyper-surface H if and only if both: to first order γ_x is

an isometry onto the tangent space of H; the mean of the curvatures of γ is equal to the mean-curvature vector of H. This gives a geometric interpretation of the drift of Brownian motion.

Note that the second condition is somewhat stronger than requiring that the image of γ has the same mean-curvature vector as H. This is because the mean-curvature vector is always orthogonal to H but the mean of the curvatures depends upon the parameterization and may have a component tangent to H.

In general an SDE on a manifold $M \subseteq \mathbb{R}^n$ can be understood as SDEs on $\mathbb{R}^n$ whose trajectories which start on M remain on M. As in the above example, at each point in the manifold M the 2-jet in $\mathbb{R}^n$ will be completely determined by the SDE on M. One can interpret the drift term of (12) for the process in $\mathbb{R}^n$ as having a tangent component determined by the intrinsic SDE on M and an orthonormal component determined by the requirement that the trajectories are confined to M.

4 Drawing SDEs driven by vector Brownian motion

We can draw an Itô SDE driven by d-dimensional Brownian motion by drawing a function

$$\gamma_x : \mathbb{R}^d \to M$$

at every point on the manifold M. Of course, in practice one only draws the function at a finite set of sample points in M.

However, γ_x is not uniquely determined by the SDE. By drawing γ_x, we are drawing a representative of the equivalence class of two jets that define the same SDE. To illustrate this, in the top line of Figure 4 we have plotted three functions $\gamma_0^* : \mathbb{R}^2 \to M$ whose 2-jets all define the same SDE. They are defined as follows.

$$\begin{aligned}
\gamma_0^A(x,y) &= x(1,0) + 2y(0,1) + 2x^2(1,0), \\
\gamma_0^B(x,y) &= x(1,0) + 2y(0,1) + 2y^2(1,0), \\
\gamma_0^C(x,y) &= x(1,0) + 2y(0,1) + (x^2+y^2)(1,0).
\end{aligned} \tag{20}$$

We have plotted the image of a circle of radius $\epsilon = 0.3$ in $\mathbb{R}^2$ under each of the maps γ_0^*. The grid lines shown are the image of polar grid lines rather than Cartesian grid lines. Polar coordinates are a more natural choice for plotting SDEs since the rotational symmetry of $\mathbb{R}^d$ corresponds to the notion of weak equivalence of SDEs.

The diffusion term of an SDE corresponds to the first order term of the jets. This is a linear mapping of the plane and hence maps the unit circle to an ellipse. This gives rise to the broadly elliptical shape of the plots. The drift term of the SDE corresponds to the mean of the image. This is marked with a star. As one can see this drift is the same for all the plots γ_0^*.

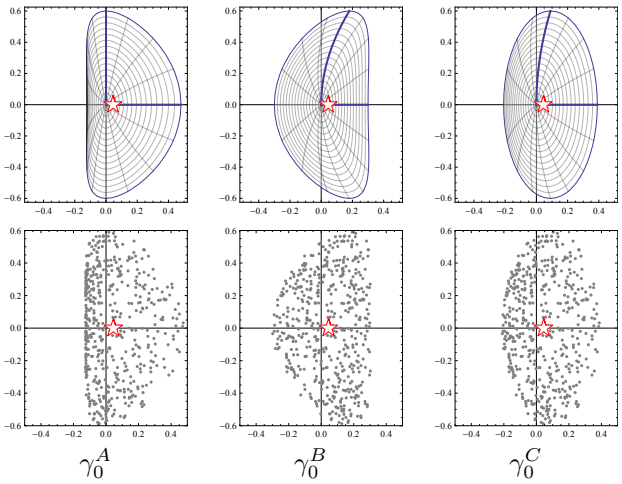

Figure 4: Plots of the equivalent 2-jets γ_0^* defined in equation 20

If we are interested in strong equivalence of the SDEs, the image of each axis is important as it tells us the strength of each component of the Brownian motion. It is only the direction of the axes that is important and not the curvature. We have used thicker lines to indicate the image of the x and y axes under each map in 4.

An alternative plot of the 2-jets γ_0^* is shown in the bottom line of Figure 4. Instead of plotting the jet by showing the image of polar grid lines, we have shown the image of a set of 1000 uniformly distributed points inside a ball of radius $\epsilon = 0.3$. We have again plotted the mean point with a star. These plots eliminate the extraneous details of the maps γ_X allowing one to see more clearly the key features that determine the weak equivalence class of the jets. These plots provide a clear visual link between the geometric and the probabilistic properties of the SDE.

If one wished to illustrate strong equivalence, the plots on the lower line in Figure 4 could be augmented with vectors indicating the mappings of each axis up to first order. Such a diagram would illustrate the key features of the strong equivalence class stripped of visual distractions such as the curvature of the images of the axes.

One can, therefore, draw a two dimensional SDE by drawing an infinitesimal diagram of the sort shown in Figure 4 at a number of points. These drawings would satisfy the desirable commutativity property illustrated in diagram (1). However, the resulting drawings would be very busy.

We can strip out some of the excessive detail from our diagram by deciding to choose a specific representative of the 2-jet at each point. Given an SDE in local coordinates,

$$\mathrm{d}X_t = a(X_t)\mathrm{d}t + b_i(X_t)\mathrm{d}W_t^i$$

we choose the specific two jet given by

$$\gamma_x(s) = x + \frac{1}{2} a\, g^E_{ij} s^i s^j + b_i s^i.$$

[DB: here I agree with the reviewer we should have $1/d$ rather than $1/2$. Take the case with d dimensional brownians and g the kronecker delta, then $g_{ij}(dW_i)(dW_j) = d\ dt$. So to get back a in the drift we need to divide by d (rather than two).

The image of an ϵ ball under γ_x will be an ellipsoid. Moreover, if we know that γ_x is of this form, we can recover the coefficients a and b_i up to weak equivalence just from knowledge of the image of the ϵ ball. As an example, notice that the curve γ_0^C from (20) is of this form. This allows us to simplify our diagrams by representing each 2-jet by drawing the image of an ϵ ball at each point under this specific representative of the 2-jet.

For example in Figure 5 we show a plot of the Heston stochastic volatility model with drift (see [22]):

$$\begin{aligned} \mathrm{d}S_t &= \mu S_t \mathrm{d}t + \sqrt{\nu_t} S_t \mathrm{d}W_t^1, \\ \mathrm{d}\nu_t &= \kappa(\theta - \nu_t)\mathrm{d}t + \xi\sqrt{\nu_t}\,(\rho \mathrm{d}W_t^1 + \sqrt{1-\rho^2}\mathrm{d}W_t^2). \end{aligned} \tag{21}$$

Note that as well as plotting the ellipses, the figure indicates the exact point that each ellipse is associated with. The extent to which the centre of the ellipse differs from the associated point is a measure of the drift.

Figure 5 is coordinate dependent since its definition depends upon choosing a specific representative of the 2-jet at each point. However, it can be thought of as a visual shorthand for a coordinate independent diagram where repeated copies of the more detailed pictures of Figure 4 are used.

Similarly Figure 6 depicts the SDE defining Brownian motion on a genus 2 surface by showing the image of the center and the 12 points on the edge of a clock-face under γ. These are shown in blue. As can be seen all the clock faces appear to be circles of the same size, this is a characteristic property one would expect of Brownian motion. It shows that the metric induced by the SDE is indeed equal to the metric induced by the embedding. Similarly the image of the centre of the clock face appears to be in the centre of each of the circles. This shows that the forward and backward operators are equal.

5 Jets, vector fields and Stratonovich calculus

We wish to show how jets can be described using vector fields. This will allow us to relate our approach to SDEs to the approach of Stratonovich calculus.

For simplicity, let us assume in this section that the driver is one dimensional. Thus to define an SDE on a manifold, one must choose a 2-jet of a curve at each point of the manifold. One way to specify a k-jet of a curve at every point in a neighbourhood is to first choose a chart for the neighbourhood and then consider curves of the form

$$\gamma_x(t) = x + \sum_{i=1}^{k} a_i(x) t^i \tag{22}$$

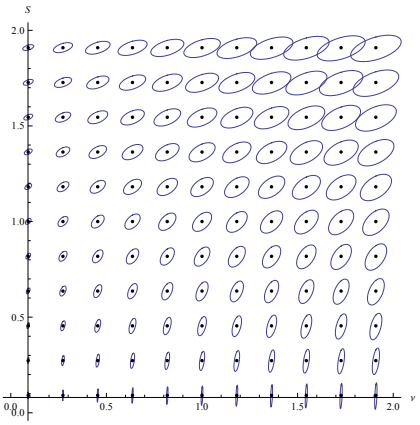

Figure 5: A plot of the Heston model (21). Parameter values $\xi = 1$, $\theta = 0.4$, $\kappa = 1$, $\mu = 0.1$, $\rho = 0.5$. We have plotted the image of the ϵ-balls for $\epsilon = 0.05$

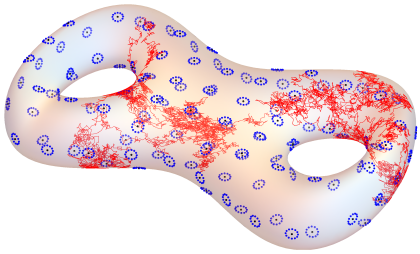

Figure 6: Plot of Brownian motion over a short, finite time interval on a genus 2 surface (red). The blue dots can be interpreted as a plot of the SDE defining Brownian motion.

where $a_i : \mathbb{R}^n \to \mathbb{R}^n$. As we have already seen in Lemma 2.1, these coefficient functions a_i depend upon the choice of chart in a relatively complex way. For example for 2-jets the coefficient functions are not vectors but instead transform according to Itô's lemma. We will call this the *standard representation* for a family of k-jets.

An alternative way to specify the k-jet of a curve at every point is to choose k vector fields $A_1, \ldots, A_k$ on the manifold. One can then define $\Phi^t_{A_i}$ to be the vector flow associated with the vector field A_i. This allows one to define curves at each point x as follows.

$$\gamma_x(t) = \Phi^{t^k}_{A_k}(\Phi^{t^{k-1}}_{A_{k-1}}(\ldots(\Phi^t_{A_1}(x))\ldots)) \tag{23}$$

where t^k denotes the k-th power of t. We will call this the *vector representation* for a family of k-jets. It is not immediately clear that all k-jets of curves can be written in this way. Let us prove this.

Suppose one chooses a chart and attempts to compute the relationship between the coefficients a_i in the standard representation and the components of the vector fields A_i in the vector representation. It is clear that the $O(t)$ term $a_1(x)$ will depend bijectively and linearly on $A_1(x)$. Thus there is a bijection between 1-jets written in the form (22) and 1-jets written in the form (23). The $O(t^2)$ term will depend linearly upon $A_2(x)$ together with a more complex term derived from A_1 and its first derivative. Symbolically $a_1(x) = \rho(A_2(x)) + f(A_1(x), (\nabla A_1)(x))$ where ρ is a linear bijection determined by the choice of chart. It follows that there is also a bijection between 2-jets written in the standard form and 2-jets written in the vector form. Inductively we have:

Theorem 5.1. *Smooth k-jets of a curves can be defined uniquely by a list of k vector fields $X_1, \ldots X_k$ according to the formula* (23)

Notice that the vector representation specifically allows us to define a *family* of k-jets varying from point to point. In more technical language, the vector representation allows us to specify a section of the bundle of k-jets. If one only specifies vectors at a point rather than vector fields, one cannot define the vector flows and so equation (23) cannot be used to define a k-jet at the point. Thus although there is no natural map from k vectors defined at a point to a k-jet of a curve, there is a natural map from k vector fields defined in a neighbourhood to a smoothly varying choice of k-jet at each point.

The standard and vector representations simply give us two different coordinate systems for the infinite dimensional space of families of k-jets.

When this general theory about k-jets is applied to stochastic differential equations one sees that two corresponding coordinate representations of a stochastic differential equation will emerge. Let us calculate in more detail correspondence between the two representations.

Lemma 5.1. *Suppose that a family of 2-jets of curves is given in the vector representation as*

$$\gamma_x(t) = \Phi_A^{t^2}(\Phi_B^t(x))$$

for vector fields A and B. Choose a coordinate chart and let A^i, B^i be the components of the vector fields in this chart. Then the corresponding standard representation for the family of 2-jets is

$$\gamma_x(t) = x + a(x)t^2 + b(x)t$$

with

$$\begin{aligned} a^i &= A^i + \frac{1}{2}\frac{\partial B^i}{\partial x^j}B^j \\ b^i &= B^i. \end{aligned}$$

Proof. By definition of the flow Φ_B^t, its components $(\Phi_B^t)^i$ satisfy

$$\frac{\partial(\Phi_B^t(x))^i}{\partial t} = B^i(\Phi_B^t(x)). \tag{24}$$

Differentiating this we have

$$\begin{aligned} \frac{\partial^2(\Phi_B^t(x))^i}{\partial t^2} &= \frac{\partial B^i}{\partial x^j}\frac{\partial(\Phi_B^t)^j}{\partial t} \\ &= B^j\frac{\partial B^i}{\partial x^j}. \end{aligned} \tag{25}$$

We now compute the derivatives of $\gamma_t(x)$. We write $(\Phi_A^{t^2})^i$ for the i-th component of the function $\Phi_A^{t^2}$.

$$\frac{\partial}{\partial t}((\Phi_A^{t^2})^i(\Phi_B^t(x))) = 2t\frac{\partial(\Phi_A^{t^2})^i}{\partial t}(\Phi_B^t) + \frac{\partial(\Phi_A^{t^2})^i}{\partial x^j}(\Phi_B^t)\frac{\partial(\Phi_B^t)^j}{\partial t}$$

Differentiating this again one obtains

$$\frac{\partial^2}{\partial t^2}((\Phi_A^{t^2})^i(\Phi_B^t(x))) = 4t^2\frac{\partial^2(\Phi_A^{t^2})^i}{\partial t^2}(\Phi_B^t) + 2\frac{\partial(\Phi_A^{t^2})^i}{\partial t}(\Phi_B^t) + 4t\frac{\partial^2(\Phi_A^{t^2})^i}{\partial x^j\partial t}(\Phi_B^t)\frac{\partial(\Phi_B^t)^j}{\partial t} + \frac{\partial^2(\Phi_A^{t^2})^i}{\partial x^j\partial x^k}(\Phi_B^t)\frac{\partial(\Phi_B^t)^j}{\partial t}\frac{\partial(\Phi_B^t)^k}{\partial t} + \frac{\partial(\Phi_A^{t^2})^i}{\partial x^j}\frac{\partial\Phi_B^t}{\partial t^2}.$$

At time $t = 0$ we have that Φ_A^0 is simply the identity. So its partial derivatives at time 0 are trivial to compute. Hence

$$\begin{aligned}\frac{\partial^2}{\partial t^2}((\Phi_A^{t^2})^i(\Phi_B^t(x)))\Big|_{t=0} &= 2\frac{\partial(\Phi_A^{t^2})^i}{\partial t}(\Phi_B^t) + \frac{\partial(\Phi_B^t)^i}{\partial t^2}\\ &= 2A^i + B^j\frac{\partial B^i}{\partial x^j}.\end{aligned}$$

The last line follows from the definition of Φ_A^t as the flow associated with the vector field A together with equation (25). We can now write down the expression for $a(x)$. The expression for $b(x)$ follows immediately from equation (24). □

What is interesting about this result is that an SDE can be defined using the coefficients a and b, which transform according to Itô's lemma, or they can be defined using vector fields A and B.

An alternative way of showing that SDEs can be defined in terms of vector fields was already known. It is given by introducing so-called Fisk–Stratonovich–McShane calculus [14] [30] [36] (Stratonovich from now on). This provides an alternative to the Itô calculus of [24]. The coefficients of SDEs written using Stratonovich calculus transform as vector fields. Indeed these coefficients are precisely the vector fields A and B we have just identified geometrically. Thus we have given a geometric interpretation of how the coordinate free notion of a 2-jet of a curve is related to the vector fields defining a Stratonovich SDE. This establishes the relationship between our jet approach to SDEs on manifolds and the more conventional approach of using Stratonovich calculus. This shows that we may view the choice between Itô or Stratonovich calculus simply as a choice of coordinates for a single underlying geometric structure. For readers not familiar with the Itô and Stratonovich stochastic calculi in a given coordinate system we refer to Appendix A of the preprint version [1].

6 Percentiles and fan diagrams

Most statistical properties of a distribution depend upon the coordinate system used. For example the definition of the mean of a process in $\mathbb{R}^n$ involves the vector space structure of $\mathbb{R}^n$. For this reason one would expect the trajectory of the mean of a process to depend upon the vector space structure. If one makes a non-linear transformation of $\mathbb{R}^n$ the trajectory of the mean changes. Indeed Itô's lemma tells us that the trajectory does not even remain constant to first order under coordinate changes.

Another manifestation of the same phenomenon is the fact that given an $\mathbb{R}$ valued random variable X and a non-linear function f, then $\mathbb{E}(f(X)) \neq f(\mathbb{E}(X))$. Again this arises because the definition of mean depends upon the vector space structure and f may not respect this.

However, the definition of the α-percentile depends only upon the ordering of $\mathbb{R}$ and not its vector space structure. As a result, for continuous monotonic f and X with connected state space, the median of $f(X)$ is equal to f applied to the median of X. If f is strictly increasing, the analogous result holds for the α percentile. If f is decreasing, the α percentile of $f(X)$ is given by f applied to the $1-\alpha$ percentile of X.

This has the implication that the trajectory of the α-percentile of an $\mathbb{R}$ valued stochastic process is invariant under smooth monotonic coordinate changes of $\mathbb{R}$. In other words, percentiles have a coordinate free interpretation. The mean does not. This raises the question of how the trajectories of the percentiles can be related to the coefficients of the stochastic differential equation. We will now calculate this relationship.

First we note that all smooth one dimensional Riemannian manifolds are isomorphic. When interpreted in terms of SDEs, this tells us that for any one dimensional SDE with non-vanishing diffusion coefficient (or volatility term)

$$\mathrm{d}X_t = a(X_t, t)\,\mathrm{d}t + b(X_t, t)\mathrm{d}W_t, \qquad X_0 = x_0, \tag{26}$$

we can find a coordinate system with respect to which the volatility term is equal to one. Such a transformation is known as a Lamperti transformation (see for example [31]) and is given by $Z_t = \phi(t, X_t)$ where $\phi(t, x)$ is a primitive integral of $1/b(x, t)$ with respect to x. Let

$$dZ_t = \alpha(Z_t, t)dt + dW_t,\ Z_0 = z_0 = \phi(0, x_0)$$

be the transformed equation. Since one dimensional Riemannian manifolds are translation invariant, we have a gauge freedom in defining a Lamperti transformation determined by the base-point of the isomorphism. Define a path $z^0(t)$ by the ordinary differential equation

$$\frac{\mathrm{d}z^0}{\mathrm{d}t} = \alpha(z^0(t), t), \quad z^0(0) = z_0.$$

If we set $Y_t = Z_t - z^0(t)$ then Y follows the SDE

$$dY_t = \bar{a}(Y_t, t)dt + dW_t, \quad \bar{a}(y, t) = \alpha(y + z^0(t), t) - \alpha(z^0(t), t), \quad Y_0 = 0, \tag{27}$$

and moreover the drift of the Y SDE vanishes at 0 for all t, $\bar{a}(0, t) = 0$.

Summing up, if $b(x, t)$ is nowhere vanishing, there is a unique time dependent transformation that maps the original SDE (26) for X into a SDE with constant diffusion coefficient, zero initial condition and zero drift at zero. Actually, sufficient conditions for the Lamperti transformed SDE (27) to have a unique strong solutions are more stringent. For example, in the autonomous

case $a(x,t) = a(x)$ and $b(x,t) = b(x)$ one may need a and b bounded from below and above, with b and its bounds strictly positive, with $a \in C^1$ and $b \in C^2$. We will assume in the following that sufficient conditions for the solution of the Lamperti transformed SDE to exist unique hold.

Thus given a one dimensional SDE with non-vanishing volatility we can always make a transformation such that (subject to bounds) the following proposition applies (we denote $\bar{a}$ by a for simplicity).

Proposition 6.1. *Let $a(x,t)$ be a bounded smooth function on $\mathbb{R} \times [0, +\infty)$ with $a(0,t) = 0$. Define the operators L and its adjoint L^* as*

$$Lp = \frac{1}{2}\frac{\partial^2 p}{\partial x^2} + a(x,t)\frac{\partial p}{\partial x} - \frac{\partial p}{\partial t}, \tag{28}$$

$$L^*p = \frac{1}{2}\frac{\partial^2 p}{\partial x^2} - \frac{\partial (a(x,t)p)}{\partial x} - \frac{\partial p}{\partial t}. \tag{29}$$

Assume further that $a(x,t)$ has additional regularity required to ensure existence and uniqueness of a fundamental solution for the PDEs $Lp = 0$ and $L^*p = 0$. Let $\Gamma(x,t;\xi,\tau)$ be the fundamental solution of $Lp = 0$. Then for $\lambda \in (0,1)$ and a fixed terminal time T, Γ satisfies

$$\Gamma(x,t;0,0) = \frac{1}{\sqrt{2\pi t}} \exp\left(-\frac{x^2}{2t}\right) + O\left(\frac{t+x^2}{\sqrt{t}} \exp\left(-\frac{\lambda x^2}{2t}\right)\right)$$

on $\mathbb{R} \times [0,T]$, and the fundamental solution of $L^*p = 0$ satisfies

$$\Gamma^*(0,0;x,t) = \Gamma(x,t;0,0)$$

(see [15] for the definition of a fundamental solution).

Proof. Equation $L^*p = 0$ as the evolution for the density of a solution of a SDE is studied for example in Friedman's SDE book [16], see Eq. 5.28 in Chapter 6: by Theorem 4.7 in Ch 6 the fundamental solution of $L^*p = 0$ is the same as the fundamental solution of $Lp = 0$. We will thus focus on $Lp = 0$, for which one can refer to Ch 1 Section 6 of Friedman's parabolic PDEs book [15]. When a is identically zero, the fundamental solution of $Lp = 0$ in (28) is given by

$$Z(x,t;\xi,\tau) = \frac{1}{\sqrt{2\pi(t-\tau)}} \exp\left(-\frac{(x-\xi)^2}{2(t-\tau)}\right).$$

By ([15] Ch 1, Theorem 10, p. 23), the fundamental solution of (28) is given by

$$\Gamma(x,t;\xi,\tau) = Z(x,t;\xi,\tau) + \int_\tau^t \int_{\mathbb{R}} Z(x,t;\eta,\sigma)\Phi(\eta,\sigma;\xi,\tau)\,\mathrm{d}\eta\,\mathrm{d}\sigma \tag{30}$$

where

$$\Phi(x,t;\xi,\tau) = \sum_{i=1}^{\infty} \Phi_i(x,t;\xi,\tau)$$

where we define $\Phi_1{=}LZ$ and for $i \geq 1$ we define

$$\Phi_{i+1}(x,t;\xi,\tau) = \int_\tau^t \int_{\mathbb{R}} L\ Z(x,t;y,\sigma)\Phi_i(y,\sigma,\xi,\tau)\,\mathrm{d}y\,\mathrm{d}\sigma.$$

Let us define:

$$\Gamma_i(x,t;\xi,\tau) := \int_\tau^t \int_{\mathbb{R}} Z(x,t;\eta,\theta)\Phi_i(\eta,\theta;\xi,\tau)\,\mathrm{d}\eta\,\mathrm{d}\theta.$$

First we bound Γ_1.

$$\begin{aligned}\Gamma_1(x,t;\xi,\tau) &= \int_\tau^t \int_{\mathbb{R}} Z(x,t;\eta,\sigma)LZ(\eta,\sigma;\xi,\tau)\,\mathrm{d}\eta\,\mathrm{d}\sigma \\ &= \int_\tau^t \int_{\mathbb{R}} Z(x,t;\eta,\sigma)a(\eta,\sigma)\frac{\partial}{\partial\eta}Z(\eta,\sigma;\xi,\tau)\,\mathrm{d}\eta\,\mathrm{d}\sigma.\end{aligned}$$

The final line follows because Z is the fundamental solution of (28) when a is equal to zero and so the parts of Γ_1 that don't involve a vanish. Our assumptions ensuring Lipschitz continuity of $a(x,t)$ uniformly in t and the fact that $a(0,t) = 0$ for all t (see (27)) tell us that there is some constant with $|a(x,t)| < C|x|$ for all x and $t \in [0,T]$. So

$$\left|a(\eta,\sigma)\frac{\partial}{\partial\eta}Z(\eta,\sigma;0,0)\right| = \left|-\frac{1}{\sqrt{2\pi\sigma}}a(\eta,\sigma)\eta\exp\left(-\frac{\eta^2}{2\sigma}\right)\right| < \frac{C}{\sqrt{2\pi\sigma}}\eta^2\exp\left(-\frac{\eta^2}{2\sigma}\right).$$

We deduce that

$$|\Gamma_1(x,t;0,0)| \leq C\int_0^t\int_{\mathbb{R}} Z(x,t;\eta,\sigma)\eta^2\frac{1}{\sqrt{2\pi\sigma}}\exp\left(-\frac{\eta^2}{2\sigma}\right)\mathrm{d}\eta\,\mathrm{d}\sigma = C\frac{e^{-\frac{x^2}{2t}}\left(t+x^2\right)}{2\sqrt{2\pi t}}.$$

The final step follows simply by the routine evaluation of the integral. To do this we use estimates which were used in [15] to prove that our expression for Γ exists and is a fundamental solution. First, we have for any $\lambda \in (0,1)$ that

$$|Z(x,t,\xi,\tau)| \leq \frac{1}{(t-\tau)^{\frac{1}{2}}}\exp\left(-\frac{\lambda|x-\xi|^2}{2(t-\tau)}\right).$$

From ([15] Ch 1, 4.14, p. 16) we have that there exist positive constants H and H_0 such that for all $\lambda \in (0,1)$ and $m \geq 2$

$$|\Phi_m(x,t;\xi,\tau)| \leq \frac{H_0H^m}{\Gamma_E(m)}(t-\tau)^{m-\frac{3}{2}}\exp\left(-\frac{\lambda|x-\xi|^2}{2(t-\tau)}\right)$$

where Γ_E is the gamma function. From ([15] Ch 1, Lemma 3 p. 15) we have that if $r < \frac{3}{2}$ and $s < \frac{3}{2}$ then

$$\begin{aligned}&\int_\sigma^\tau\int_{\mathbb{R}}(t-\tau)^{-r}\exp\left(-\frac{h(x-\xi)^2}{2(t-\tau)}\right)(\tau-\sigma)^{-s}\exp\left(-\frac{h(x-\xi)^2}{2(t-\sigma)}\right)\mathrm{d}\xi\,\mathrm{d}\sigma \\ &\qquad = \sqrt{\frac{2\pi}{h}}B\left(\frac{3}{2}-r,\frac{3}{2}-s\right)(t-\sigma)^{\frac{3}{2}-r-s}\exp\left(-\frac{h(x-y)^2}{2(t-\sigma)}\right)\end{aligned}$$

where B is the beta function. Taking $r = 1/2$ and $s = \frac{3}{2} - m$ we obtain the estimate

$$\begin{aligned}|\Gamma_m(x,t;\xi,\tau)| &\leq \frac{H_0 H^m}{\Gamma_E(m)}\sqrt{\frac{2\pi}{\lambda}}B(1,m)(t-\tau)^{-\frac{1}{2}+m}\exp\left(-\frac{\lambda(x-\xi)^2}{2(t-\tau)}\right) \\ &= \frac{H_0 H^m}{m!}\sqrt{\frac{2\pi}{\lambda}}(t-\tau)^{-\frac{1}{2}+m}\exp\left(-\frac{\lambda(x-\xi)^2}{2(t-\tau)}\right)\end{aligned}.$$

So

$$\begin{aligned}\sum_{k=2}^{\infty}|\Gamma_m(x,t;0,0)| &\leq \sum_{m=2}^{\infty} H_0\frac{H^m}{m!}t^{-\frac{1}{2}+m}\exp\left(-\frac{\lambda x^2}{2t}\right) \\ &= \sqrt{\frac{2\pi}{\lambda}}H_0 H^2\exp(Ht)(t)^{3/2}\exp\left(-\frac{\lambda x^2}{2t}\right)\end{aligned}$$

The result follows if we conclude by invoking Theorem 15, Ch 1, p. 28 of [15], or Theorem 4.7 in Ch 6 of [16]. □

Theorem 6.1. *For sufficiently small t, the α-th percentile of the solutions to* (26) *is given by*

$$x_0 + b_0\sqrt{t}\Phi^{-1}(\alpha) + \left(a_0 - \frac{1}{2}b_0 b_0'(1-\Phi^{-1}(\alpha)^2)\right)t + O(t^{3/2}) \tag{31}$$

so long as the coefficients of (26) *are smooth, the diffusion coefficient b never vanishes, and sufficient conditions for the Lamperti transformed SDE and for $L^*p = 0$ to have a unique regular solution hold. In this formula a_0 and b_0 denote the values of $a(x_0,0)$ and $b(x_0,0)$ respectively. In particular, the median process is a straight line up to $O(t^{\frac{3}{2}})$ with tangent given by the drift of the Stratonovich version of the Itô SDE* (26)*. The $\Phi(1)$ and $\Phi(-1)$ percentiles correspond up to $O(t^{\frac{3}{2}})$ to the curves $\gamma_{X_0}(\pm\sqrt{t})$ where γ_{X_0} is any representative of the 2-jet that defines the SDE in Itô form.*

Proof. We first apply a Lamperti transformation so that the conditions of Proposition 6.1 apply. Let write y for the coordinates after applying the Lamperti transformation and let us write ρ for the 1-form representing the probability measure. By Proposition 6.1

$$\rho = \left[\frac{1}{\sqrt{2\pi t}}\exp\left(-\frac{y^2}{2t}\right) + O\left(\frac{t+y^2}{\sqrt{t}}\exp\left(-\frac{\lambda y^2}{2t}\right)\right)\right]\mathrm{d}y.$$

We introduce a new coordinate $z = \frac{y}{\sqrt{t}}$ so that the ρ can be written

$$\rho = \left[\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{z^2}{2}\right) + O\left(t(1+z^2)\exp\left(-\frac{\lambda z^2}{2}\right)\right)\right]\mathrm{d}z.$$

Integrating this, we see that for sufficiently small ϵ we have a uniform estimate

$$\int_{-\infty}^{\Phi^{-1}(\alpha+\epsilon)}\rho = \alpha + \epsilon + O(t).$$

It follows that the α-th percentile in z coordinates is $\Phi^{-1}(\alpha) + O(t)$. Hence in y coordinates it is $\phi^{-1}(\alpha)\sqrt{t} + O(t^{3/2})$.

Let g denote the inverse of the Lamperti transformation that we have taken. Let us write the Taylor series expansion for g around (0,0) as

$$g(y,t) = x_0 + g_y y + \frac{1}{2} g_{yy} y^2 + g_t t + O(y^3 + t^2). \tag{32}$$

By construction the SDE in y coordinates is of the form

$$\mathrm{d}Y_t = \overline{a}\mathrm{d}t + \mathrm{d}W_t$$

with $\overline{a}(0,t) = 0$. When we apply g to this equation we will get equation (26). Using Itô's lemma we deduce that:

$$b = g_y, \quad a_0 = g_t + \frac{1}{2} g_{yy}.$$

The first of these equations holds everywhere, the second uses the vanishing of $\overline{a}$ at 0. To avoid notational ambiguity in partial differentiation, we will temporarily write s for the time coordinate when paired with x and t for the time coordinate when paired with y. Thus we are making the coordinate transformation

$$x = g(y,t), \quad s = t.$$

We then have that

$$\frac{\partial b}{\partial x} = \frac{\partial b}{\partial y}\frac{\partial y}{\partial x} + \frac{\partial b}{\partial s}\frac{\partial s}{\partial x} = g_{yy}\frac{1}{g_y}.$$

So that we have that at 0

$$g_y = b_0, \quad g_{yy} = b_0 b_0', \quad g_t = \left(a_0 - \frac{1}{2} b_0 b_0'\right).$$

Substituting these formulae into the Taylor series expansion (32) for g yields

$$g(y,t) = x_0 + b_0 y + \frac{1}{2} b_0 b_0' y^2 + (a_0 - \frac{1}{2} b_0 b_0')t + O(y^3 + t^2).$$

Substitute $y = \Phi^{-1}(\alpha)\sqrt{t}$ to get

$$x_0 + b_0 \Phi^{-1}(\alpha)\sqrt{t} + \frac{1}{2} b_0 b_0' \Phi^{-1}(\alpha)^2 t + (a_0 - \frac{1}{2} b_0 b_0')t + O(t^{3/2}).$$

This simplifies to (31).

□

The theorem above has given us the median as a special case, and a link between the median and the Stratonovich version of the SDE. By contrast the mean process has tangent given by the drift of the Itô SDE as the Itô integral is a martingale.

For completeness, besides the mean and the median, we wish to consider the mode. We claim that under the same conditions as the theorem above, there are paths $m^u(t)$ and $m^l(t)$ both satisfying

$$m^u = x_0 + a(x_0, 0)t - \frac{3}{2}b(x_0, 0)b'(x_0, 0)t + O(t^{\frac{3}{2}})$$
$$m^l = x_0 + a(x_0, 0)t - \frac{3}{2}b(x_0, 0)b'(x_0, 0)t + O(t^{\frac{3}{2}})$$

such that for sufficiently small t there exists a mode lying in $[m^l(t), m^u(t)]$ This relationship between the mean, median and mode is approximately seen in many general probability distributions as was observed qualitatively by Pearson [32].

This result gives an alternative way of plotting the two jet that defines a one dimensional stochastic differential equation in terms of fan diagrams. A *fan diagram* is a standard tool in econometrics for illustrating the predictions of a model. In Figure 7 we have plotted a fan diagram for a stock price modelled by geometric Brownian motion. The negative times in the plot show historical values for the stock price. For positive times, we plot a single random realization of geometric Brownian motion together with two percentiles that indicate the range of values attained by other realizations. We have chosen to plot the percentiles $\Phi(1) \approx 84\%$ and $\Phi(-1) \approx 16\%$.

We can use the result above to plot an SDE by drawing an infinitesimal fan-diagram at each point. At each point $x \in \mathbb{R}$ one plots the curve $(t, \gamma_x(\pm\sqrt{t}))$. One interprets this diagram as an infinitesimal fan-diagram showing the $\Phi(1)$ and $\Phi(-1)$ percentiles. Such a plot is shown for the process $\mathrm{d}S_t = \sigma S_t \mathrm{d}W_t$ in the left hand panel of 8.

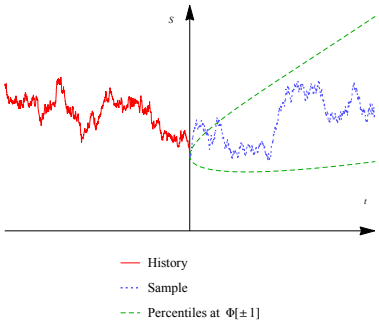

Figure 7: A fan diagram for a stock price, S modelled using geometric Brownian motion

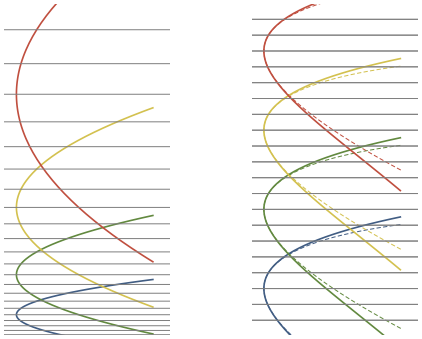

Figure 8: A fan diagram of $\mathrm{d}S_t = S_t\sigma_t \mathrm{d}W_t$ (left). The result of applying $(x, y) \to (x, \log y)$ to this fan diagram (right, solid line). A fan diagram of equation (33) (right, dashed line).

In the right hand panel of 8 we show how this plot transforms when one sends (t, S_t) to $(t, \log(S_t))$. This is illustrated with solid lines. We also use

dashed lines to plot the corresponding diagram for the equation arising from Itô's lemma, namely

$$\mathrm{d}(\log(S_t)) = -\frac{1}{2}\sigma^2 \mathrm{d}t + \sigma \, \mathrm{d}W_t. \tag{33}$$

It is interesting to notice that one can clearly see the drift term in the right hand side of 8. Notice also that this drift arises because the spacing between the grid lines on the left hand side of 8 increases as one moves up the page whereas the corresponding grid lines after transformation are evenly spaced. This can be interpreted as a visual demonstration that the Itô term in the transformation rule for SDEs is determined by the second derivative of the transformation.

7 Conclusions and further work

In this first work we showed how Itô SDEs driven by Brownian motion could be understood in terms of 2-jets of maps. In further work we will study more in detail the relationship between the jets approach and Schwartz morphism based on second order tangent vectors and co-vectors (as in the Schwartz - Meyer theory explained in Emery's work [12]). We will study the relationship with Belopolskaja and Dalecky's Itô bundle [5], see also [20] and the appendix in [7]. We will explore the jet approach in connection with projection method for dimensionality reduction and approximation of SDEs [2, 3]. We will explore practical applications of our apporach to simulating SDEs on manifolds [4]. We plan to investigate the use of jets in rough paths theory [17].

A Proof of convergence to classical Itô calculus

In this appendix we prove Theorem 2.1, stating convergence in $L^2(\mathbb{P})$ of the 2-jet scheme to the classical Itô solution of the SDE. The techniques used to prove almost-sure convergence of the classical Euler scheme ([21] for example) could be adapted to the 2-jet scheme.

Proof. We think of T as fixed and as N increases $\mathcal{T}^N$ provides a finer discretization grid approximating $[0, T]$.

To remove clutter from our equations, during this proof we will adopt the following conventions. C is a constant, independent of N that may change from line to line. We drop the superscript δt from terms such as $X_t^{\delta t}$. Sums over Greek indices always range from 1 to d. i, j and k are always integers. Terms with integer time subscripts such as X_i are shorthand for $X_{i\delta t}$. Superscript T's indicate the vector transpose rather than the terminal time.

Under our hypotheses, we know from [27], Theorem 10.2.2 p. 342 that the Euler scheme

$$\bar{X}_{t+\delta t} = \bar{X}_t + a(\bar{X}_t)\delta t + \sum_\alpha b_\alpha(\bar{X}_t)(W^\alpha_{t+\delta t} - W^\alpha_t) \text{ for } \ t \in \mathcal{T}^N$$

converges to the solution of the Itô SDE in that

$$\max_{t\in\mathcal{T}^N} E\left\{|\bar{X}_t - \tilde{X}_t|^2\right\} \leq C\ \delta t\ .$$

Since

$$\begin{aligned}\max_{t\in\mathcal{T}^N} E\left\{|X_t - \tilde{X}_t|^2\right\} &\leq \max_{t\in\mathcal{T}^N} E\left\{2|X_t - \bar{X}_t|^2 + 2|\bar{X}_t - \tilde{X}_t|^2\right\} \\ &\leq 2\max_{t\in\mathcal{T}^N} E\left\{|X_t - \bar{X}_t|^2\right\} + 2\max_{t\in\mathcal{T}^N} E\left\{|\bar{X}_t - \tilde{X}_t|^2\right\}\end{aligned}$$

all we need to conclude with (15) for time points in $\mathcal{T}^N$ is to show

$$\max_{t\in\mathcal{T}^N} E\left\{|X_t - \bar{X}_t|^2\right\} \leq C\ \delta t.$$

Summing the differences of consecutive terms in the Euler scheme with time step δt we have

$$\bar{X}_k = x_0 + \sum_{i=1}^{k}\left[a(\bar{X}_{i-1})\delta t + \sum_\alpha b_\alpha(\bar{X}_{i-1})\delta W_{i-1}^\alpha\right]$$

where $\delta W_k := W_{(k+1)\delta t} - W_{k\delta t}$. Using the definition for a we can write this as

$$\bar{X}_k = x_0 + \sum_{i=1}^{k}\left[\sum_{\alpha,\beta}\tilde{a}_{\alpha,\beta}(\bar{X}_{i-1})g_E^{\alpha\beta}\delta t + \sum_\alpha b_\alpha(\bar{X}_{i-1})\delta W_{i-1}^\alpha\right] \tag{34}$$

where g_E is the metric tensor of the Euclidean metric and where we define

$$\tilde{a}_{\alpha,\beta}(x) := \frac{1}{2}\frac{\partial^2\gamma_x}{\partial u^\alpha \partial u^\beta}\Big|_{u=0}, \quad \text{so that} \quad a = \sum_{\alpha,\beta}\tilde{a}_{\alpha,\beta}g^{\alpha,\beta}.$$

Note that we are not using the Einstein summation convention in this proof. Let us write the 2-jet scheme via its Taylor expansion as in (4), namely

$$X_k = X_{k-1} + \sum_{\alpha,\beta}\tilde{a}_{\alpha,\beta}(X_{k-1})\delta W_{k-1}^\alpha \delta W_{k-1}^\beta + \sum_\alpha b_\alpha(X_{k-1})\delta W_{k-1}^\alpha + R_k. \tag{35}$$

This expression defines the remainder R_k. Note that the remainder depends also on N but our notation has suppressed this. Summing the differences of consecutive terms, and substituting in the definition of b we obtain

$$X_k = x_0 + \sum_{i=1}^{k}\left[\sum_{\alpha,\beta}\tilde{a}_{\alpha,\beta}(X_{i-1})\delta W_{i-1}^\alpha \delta W_{i-1}^\beta + \sum_\alpha b_\alpha(X_{i-1})\delta W_{i-1}^\alpha + R_i\right]. \tag{36}$$

Subtracting equation (34) from (36) we obtain

$$X_k - \bar{X}_k = P_k + \sum_\alpha Q_k^\alpha + \sum_{\alpha,\beta} S_k^{\alpha,\beta}$$

where we define

$$P_k := \sum_{i=1}^{k} R_i, \qquad Q_k^\alpha := \sum_{i=1}^{k} \left[b_\alpha(X_{i-1}) - b_\alpha(\bar{X}_{i-1}) \right] \delta W_{i-1}^\alpha$$

$$\text{and } S_k^{\alpha,\beta} := \sum_{i=1}^{k} \left[\tilde{a}_{\alpha,\beta}(X_{i-1}) \delta W_{i-1}^\alpha \delta W_{i-1}^\beta - \tilde{a}_{\alpha,\beta}(\bar{X}_{i-1}) g_E^{\alpha\beta} \delta t \right].$$

We have that

$$E\left\{ |X_k - \bar{X}_k|^2 \right\} \leq C \left(E\left\{ |P_k|^2 \right\} + \sum_\alpha E\left\{ |Q_k^\alpha|^2 \right\} + \sum_{\alpha,\beta} E\left\{ |S_k^{\alpha,\beta}|^2 \right\} \right).$$

We now wish to obtain bounds for each expectation on the right in terms of δt and the function.

$$Z(t) = \max_{0 \leq s \leq t, s \in \mathcal{T}^N} E\left\{ |X_s - \bar{X}_s|^2 \right\}.$$

Using our bound on the third derivatives of γ, we can bound the remainder terms R_i as follows

$$|R_i| \leq C \Big(\sum_\alpha |\delta W_{i-1}^\alpha| \Big)^3.$$

Writing $M := E\Big\{ \Big(\sum_\alpha |\delta W_{i-1}^\alpha| \Big)^6 \Big\}$, we calculate that

$$E(|P_k|^2) = E\{ |\sum_i R_i|^2 \} \leq CN^2 M \leq C \left(\frac{T}{\delta t} \right)^2 (\delta t)^{\frac{6}{2}} \leq C \delta t.$$

By the discrete Itô isometry and the Lipschitz condition on the derivatives we find

[DB: here I found a missing δt]

$$E\Big\{ |Q_k^\alpha|^2 \Big\} = E\Big\{ \Big| \sum_{i=1}^{k} [b_\alpha(X_{i-1}) - b_\alpha(\bar{X}_{i-1})] \delta W_{i-1}^\alpha \Big|^2 \Big\} = E\Big\{ \Big| \sum_{i=1}^{k} [b_\alpha(X_{i-1}) - b_\alpha(\bar{X}_{i-1})]^2 \Big| \delta t \Big\}$$
$$\leq CE\Big\{ \Big| \sum_{i=1}^{k} |X_{i-1} - \bar{X}_{i-1}| \delta W_{i-1}^\alpha \Big|^2 \Big\} \leq C \int_0^{k\delta t} Z(s)\, \mathrm{d}s.$$

To bound $S^{\alpha,\beta}$ we write it as a sum of two components $S^{\alpha,\beta,1}$ and $S^{\alpha,\beta,2}$ defined as follows.

$$S_k^{\alpha,\beta,1} := \sum_{i=1}^{k} \tilde{a}_{\alpha,\beta}(\bar{X}_{i-1}) (\delta W_{i-1}^\alpha \delta W_{i-1}^\beta - g_E^{\alpha,\beta} \delta t)$$

$$S_k^{\alpha,\beta,2} := \sum_{i=1}^{k} [\tilde{a}_{\alpha,\beta}(X_{i-1}) - \tilde{a}_{\alpha,\beta}(\bar{X}_{i-1})] \delta W_{i-1}^\alpha \delta W_{i-1}^\beta$$

By expanding the square term using the definition of $S^{\alpha,\beta,1}$ we write $E\left\{|S_k^{\alpha,\beta,1}|^2\right\}$ as

$$\sum_{i=1}^{k}\sum_{j=1}^{k} E\left\{\tilde{a}^T_{\alpha,\beta}(\bar{X}_{i-1})\tilde{a}_{\alpha,\beta}(\bar{X}_{j-1})(\delta W^\alpha_{i-1}\delta W^\beta_{i-1} - g_E^{\alpha,\beta}\delta t)(\delta W^\alpha_{j-1}\delta W^\beta_{j-1} - g_E^{\alpha,\beta}\delta t)\right\}.$$

When $i \neq j$ the terms on the right hand side vanish. This is because we may assume WLOG that $j > i$ in which case the last factor of the (i,j)-th term,

$$\delta W^\alpha_{j-1}\delta W^\beta_{j-1} - g_E^{\alpha,\beta}\delta t$$

is independent of the rest of the term and has expectation 0. We now quote 10.2.14 p. 343 in [27] to show

$$E\left\{|\bar{X}_k|^2\right\} \leq C. \tag{37}$$

We now compute

$$\begin{aligned}
E\left\{|S_k^{\alpha,\beta,1}|^2\right\} &= \sum_{i=1}^{k} E\left\{|\tilde{a}_{\alpha,\beta}(\bar{X}_{i-1})|^2(\delta W^\alpha_{i-1}\delta W^\beta_{i-1} - g_E^{\alpha,\beta}\delta t)^2\right\} \\
&\leq \sum_{i=1}^{k} E\left\{|\tilde{a}_{\alpha,\beta}(\bar{X}_{i-1})|^2\right\} E\left\{(\delta W^\alpha_{i-1}\delta W^\beta_{i-1} - g_E^{\alpha,\beta}\delta t)^2\right\} \\
&\leq \sum_{i=1}^{k} C(\delta t)^2 E\left\{|\tilde{a}_{\alpha,\beta}(\bar{X}_{i-1})|^2\right\} \\
&\leq \sum_{i=1}^{k} C(\delta t)^2 E\left\{(1+\bar{X}^2_{i-1})\right\} \leq \sum_{i=1}^{k} C(\delta t)^2 \leq C\delta t.
\end{aligned}$$

The second line follows from independence of Brownian motion increments. The third from the second by the scaling properties of Brownian motion increments. The fourth from our growth bounds on the second derivatives. We then use the bound (37). Let us write $\Delta_{\alpha,\beta,i} := \tilde{a}_{\alpha,\beta}(X_i) - \tilde{a}_{\alpha,\beta}\bar{X}_i$. We find

$$\begin{aligned}
E\left\{|S_k^{\alpha,\beta,2}|^2\right\} &= 2\sum_{i=0}^{k-1}\sum_{j=0}^{i} E\left\{\Delta^T_{\alpha,\beta,i}\Delta_{\alpha,\beta,j}\delta W^\alpha_i \delta W^\beta_i \delta W^\alpha_j \delta W^\beta_j\right\} \\
&= 2\sum_{i=0}^{k-1}\sum_{j=0}^{i} E\left\{\Delta^T_{\alpha,\beta,i}\Delta_{\alpha,\beta,j}\delta W^\alpha_j \delta W^\beta_j\right\} E\left\{\delta W^\alpha_i \delta W^\beta_i\right\} \\
&\leq 2\sum_{i=0}^{k-1}\sum_{j=0}^{i} E\left\{|\Delta^T_{\alpha,\beta,i}|^2\right\}^{\frac{1}{2}} E\left\{|\Delta^T_{\alpha,\beta,j}\delta W^\alpha_j \delta W^\beta_j|^2\right\}^{\frac{1}{2}} E\left\{\delta W^\alpha_i \delta W^\beta_i\right\} \\
&\leq 2\sum_{i=0}^{k-1}\sum_{j=0}^{i} E\left\{|\Delta_{\alpha,\beta,i}|^2\right\}^{\frac{1}{2}} E\left\{|\Delta_{\alpha,\beta,j}|^2\right\}^{\frac{1}{2}} E\left\{|\delta W^\alpha_j \delta W^\beta_j|^2\right\}^{\frac{1}{2}} E\left\{\delta W^\alpha_i \delta W^\beta_i\right\} \\
&\leq C(\delta t)^2 \sum_{i=0}^{k-1}\sum_{j=0}^{i} \left(E\left\{|\Delta_{\alpha,\beta,i}|^2\right\} + E\left\{|\Delta_{\alpha,\beta,j}|^2\right\}\right).
\end{aligned}$$

[DB: It seems to me this above does work apart from $\sqrt{3}$ but is redundant since $E\left\{\delta W_i^\alpha \delta W_i^\beta\right\}$ can be either 0 for $\alpha \neq \beta$ or δt in the other case. We should say this. Also, there is an error for $i = j$ since in that case we cannot factor as you do, although we get a completely analogous result. So I would do the proof assuming $i < j$, saying that the case $i = j$ can be developed along the same lines, and I would make the cases $\alpha \neq \beta$ and $\beta = \alpha$ explicit. In the latter case the quantity $E\left\{\delta W_i^\alpha \delta W_i^\alpha\right\} = \delta t$ and then we proceed with Cauchy Schwartz. So it would go like this:

$$E\left\{|S_k^{\alpha,\beta,2}|^2\right\} =$$

$$= 2\sum_{i=0}^{k-1}\sum_{j=0}^{i} E\left\{\Delta^T_{\alpha,\beta,i}\Delta_{\alpha,\beta,j}\delta W_i^\alpha \delta W_i^\beta \delta W_j^\alpha \delta W_j^\beta\right\} = \text{(by independence)} =$$

$$= 2\sum_{i=0}^{k-1}\sum_{j=0}^{i} E\left\{\Delta^T_{\alpha,\beta,i}\Delta_{\alpha,\beta,j}\delta W_j^\alpha \delta W_j^\beta\right\} E\left\{\delta W_i^\alpha \delta W_i^\beta\right\} = 0 \text{ if } \alpha \neq \beta.$$

$$\text{for } \alpha = \beta \text{ by Cauchy Schwarz } \leq 2\sum_{i=0}^{k-1}\sum_{j=0}^{i} E\left\{|\Delta^T_{\alpha,\alpha,i}|^2\right\}^{\frac{1}{2}} E\left\{|\Delta^T_{\alpha,\alpha,j}\delta W_j^\alpha \delta W_j^\alpha|^2\right\}^{\frac{1}{2}} \delta t$$

$$\text{by independence} = 2\sum_{i=0}^{k-1}\sum_{j=0}^{i} E\left\{|\Delta_{\alpha,\alpha,i}|^2\right\}^{\frac{1}{2}} E\left\{|\Delta_{\alpha,\alpha,j}|^2\right\}^{\frac{1}{2}} \sqrt{3}\ (\delta t)^2$$

$$\text{by Young inequality } \leq C(\delta t)^2 \sum_{i=0}^{k-1}\sum_{j=0}^{i} \left(E\left\{|\Delta_{\alpha,\alpha,i}|^2\right\} + E\left\{|\Delta_{\alpha,\alpha,j}|^2\right\}\right).$$

[DB: Using schwartz here you get $E[(\delta W)^4] = 3\delta t$ from which the square root of 3 comes] [DB: as I was saying above the case $i = j$ would factor directly a $E[|\Delta|^2(\delta W)^4] = 3(\delta t)^2 E[|\Delta|^2]$ and is ok] We have assumed $i < j$, but the case $i = j$ is simpler and is dealt with along the same lines . [DB: this would go as steps are now in the equation:](We have used independence, Cauchy-Schwarz, independence, the scaling of Brownian motion and Young's Inequality in that order.)

Applying now the Lipschitz property of $\tilde{a}_{\alpha,\beta}$ we find

$$E\left\{|S_k^{\alpha,\beta,2}|^2\right\} \leq C(\delta t)^2 \sum_{i=0}^{k-1}\sum_{j=0}^{i} \left(E\left\{|X_i - \bar{X}_i|^2\right\} + E\left\{|X_j - \bar{X}_j|^2\right\}\right)$$

$$\leq C\delta t \sum_{i=0}^{k-1} Z(i\delta t) \leq C\int_0^{k\delta t} Z(t)\ dt$$

where we took into account the fact tha $i \leq N = T/\delta t$, so that $i(\delta t)^2 \leq C\delta t$. Putting together all our bounds we have

$$Z(t) \leq C\left(\delta t + \int_0^t Z(s)\,\mathrm{d}s\right).$$

So by the Gronwall inequality, $Z(t) \leq C\delta t$.

We have established that

$$\max_{t \in \mathcal{T}^N} E\left\{|X_t^N - \tilde{X}_t|^2\right\} \leq C\ \delta t. \tag{38}$$

To complete the proof of (15), let $t \in [0,T]$ be a time point not necessarily in the grid. Using Theorem 4.5.4 of [27] we can obtain a bound $E(|\tilde{X}_{t_i}|^2) \leq C$ and hence a bound $E(|X_{t_i}|^2) \leq C$. Applying Theorem 4.5.4 of [27] a second time we have

$$E(|\tilde{X}_t - \tilde{X}_{t_i}|^2) \leq C\left(1 + E(|\tilde{X}_{t_i}|^2)\right)\delta t \leq C\delta t.$$

By the definition of our scheme, our estimates on the derivatives of γ and Taylor's theorem

$$E(|X_t - X_{t_i}|^2) \leq C\left(1 + E(|X_{t_i}|^2)\right)\delta t \leq C\delta t.$$

The inequality (15) now follows.

□

B A coordinate free notion of convergence

The notion of mean square convergence depends upon one's choice of coordinates and so is not the best notion of convergence to use in differential geometry. Let us briefly describe an alternative form of convergence which can be used instead.

Let M be a manifold and g be a Riemannian metric on M. Let K be a compact subset M. Let d^g denote the Riemannian distance function. Let K^0 denote the interior of K. We define an equivalence relation $\sim$ on M by $x \sim y$ if either $x = y$ or both $x \notin K^0$ and $y \notin K^0$. The quotient space $M/\sim$ is simply the one-point compactification of K^0. We write ∞ for the equivalence class consisting of all points outside K^0. We may define a semimetric $\tilde{d}^{g,K}$ on $M/\sim$ by

$$\tilde{d}^{g,K}([x],[y]) = \inf_{X \sim x, Y \sim y} d^g(X,Y).$$

This is not a metric since $\tilde{d}^{g,K}$ does not obey the triangle inequality. Nevertheless, convergence of a sequence in $\tilde{d}^{g,K}$ implies convergence in $M/\sim$

Given a stochastic process $X : [0,T] \to M \cup \{\infty\}$ and a compact subset K of M we define a new stochastic process X^K by

$$X_t^K(\omega) = \begin{cases} X_t(\omega) & \text{if } X_{t'}(\omega) \in K^0 \text{ for all } t' < t \\ \infty & \text{otherwise.} \end{cases}.$$

Definition B.1. Let X^i be a sequence of stochastic processes in $M \cup \{\infty\}$. For a fixed time t, we say that X^i converges to X *in mean square on compacts* if for all $\epsilon > 0$, compact sets $K \subseteq M$ and Riemannian metrics g on M there exists $N \in \mathbb{N}$ such that if $i \geq N$

$$E(\tilde{d}^{g,K}((X^i)_t^K(\omega), X_t^K(\omega))^2) \leq \epsilon$$

If the coefficients of an SDE are smooth then all bounds required in proving Theorem 2.1 will automatically hold over compact sets. As shown in [23] any smooth Ito SDE on a manifold has a unique solution in $M \cup \{\infty\}$ if one sets the value of the solution to ∞ at the explosion time. Our 2-jet scheme converges in mean square on compacts to this solution of the corresponding Itô SDE

References

[1] J. Armstrong and D. Brigo. Coordinate-free stochastic differential equations as jets. *Arxiv preprint*. `http://arxiv.org/abs/1602.03931`.

[2] J. Armstrong and D. Brigo. Optimal approximation of SDEs on submanifolds: the ito-vector and ito-jet projections. *Arxiv preprint*. `https://arxiv.org/abs/1610.03887`.

[3] J. Armstrong and D. Brigo. Extrinsic projection of Itô SDEs on submanifolds with applications to non-linear filtering. In F. Nielsen, F. Critchley, and K. Dodson, editors, *Computational Information Geometry for Image and Signal Processing*, pages 101–120. Springer, 2017.

[4] J. Armstrong and A. Mijatovic. Simulating SDEs on manifolds. *KCL Pure working paper*, 2017. `https://tinyurl.com/ydenra62`.

[5] Y. Belopolskaja and Y. Dalecky. *Stochastic Equations and Differential Geometry*. Mathematics and Its Applications, Vol. 30. Dordrecht, Kluwer Academic Publishers, Boston, London, 1990.

[6] M. Brubaker, M. Salzmann, and R. Urtasun. A family of MCMC methods on implicitly defined manifolds. In *Artificial Intelligence and Statistics*, pages 161–172, 2012.

[7] Z. Brzeźniak and K. D. Elworthy. Stochastic differential equations on Banach manifolds. *Methods Funct. Anal. Topology*, 6(1):43–84, 2000.

[8] S. Byrne and M. Girolami. Geodesic Monte Carlo on embedded manifolds. *Scandinavian Journal of Statistics*, 40(4):825–845, 2013.

[9] P. Diaconis, S. Holmes, M. Shahshahani, et al. Sampling from a manifold. In *Advances in Modern Statistical Theory and Applications: A Festschrift in honor of Morris L. Eaton*, pages 102–125. Institute of Mathematical Statistics, 2013.

[10] K. D. Elworthy. *Stochastic differential equations on manifolds*. Cambridge University Press, Cambridge, New York, 1982.

[11] K. D. Elworthy. Geometric aspects of diffusions on manifolds. In *École d'Été de Probabilités de Saint-Flour XV–XVII, 1985–87*, pages 277–425. Springer, 1988.

[12] M. Emery. *Stochastic calculus in manifolds*. Springer-Verlag, Heidelberg, 1989.

[13] M. Emery. An invitation to second-order stochastic differential geometry. *HAL research report <hal-00145073>*, 2007.

[14] D. Fisk. Quasi-martingales and stochastic integrals. *PhD Dissertation, Michigan State University, Department of Statistics*, 1963.

[15] A. Friedman. *Partial differential equations of parabolic type*. Prentice-Hall, reprinted in 2008 by Dover-Publications, Englewood Cliffs N.J, 1964.

[16] A. Friedman. *Stochastic differential equations and applications, vol. I and II*. Academic Press, reprinted by Dover Publications, New York, 1975.

[17] P. Friz and M. Hairer. *A Course on Rough Paths with an introduction to regularity structures*. Springer-Verlag, Heidelberg, 2014.

[18] R. Gangolli. On the construction of certain diffusions on a differentiable manifold. *Probability Theory and Related Fields*, 2(5):406–419, 1964.

[19] M. Girolami and B. Calderhead. Riemann manifold Langevin and Hamiltonian Monte Carlo methods. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(2):123–214, 2011.

[20] Y. Gliklikh. *Global and Stochastic Analysis with Applications to Mathematical Physics*. Theoretical and Mathematical Physics. Springer, London, 2011.

[21] I. Gyöngy. A note on Euler's approximations. *Potential Analysis*, (8):205–216, 1998.

[22] S. Heston. A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Review of financial studies*, 6(2):327–343, 1993.

[23] Elton P Hsu. *Stochastic analysis on manifolds*, volume 38. American Mathematical Soc., 2002.

[24] K. Itô. Stochastic Integral. *Proceedings of the Imperial Academy*, 20(8):519–524, 1944.

[25] K. Itô. Stochastic differential equations in a differentiable manifold. *Nagoya Math. J.*, (1):35–47, 1950.

[26] J. Jost. *Riemannian geometry and geometric analysis*. Springer Science & Business Media, 2008.

[27] P. Kloeden and E. Platen. *Numerical solution of stochastic differential equations*. Applications of mathematics. Springer, Berlin, New York, Third printing, 1999.

[28] P. E. Kloeden and E. Platen. Higher-order implicit strong numerical schemes for stochastic differential equations. *Journal of statistical physics*, 66(1-2):283–314, 1992.

[29] S. Kobayashi and K. Nomizu. *Foundations of differential geometry*, volume 1. New York, 1963.

[30] E. J. McShane. *Stochastic Calculus and Stochastic Models (2nd Ed.)*. Academic Press, New York, 1974.

[31] G. A. Pavliotis. *Stochastic Processes and Applications: Diffusion Processes, the Fokker-Planck and Langevin Equations*. Springer, Heidelberg, 2014.

[32] K. Pearson. Contributions to the mathematical theory of evolution. ii. skew variation in homogeneous material. *Philosophical Transactions of the Royal Society of London. A*, 186:343–414, 1895.

[33] M A Pinsky. Isotropic transport process on a riemannian manifold. *Transactions of the American Mathematical Society*, 218:353–360, 1976.

[34] L. Rogers and D. Williams. Diffusions, Markov processes and martingales, vol 2: Ito calculus, 1987.

[35] D J Saunders. *The geometry of jet bundles*, volume 142 of *London Mathematical Society Lecture Note Series*. Cambridge University Press, 1989.

[36] R. L. Stratonovich. A new representation for stochastic integrals and equations. *SIAM Journal on Control*, 4(2):362–371, 1966.